Review

Mass Spectrometric Methods for Non-Targeted Screening of Metabolites: A Future Perspective for the Identification of Unknown Compounds in Plant Extracts

Michael Sasse [ID] and Matthias Rainer *[ID]

Institute of Analytical Chemistry and Radiochemistry, CCB-Center of Chemistry and Biomedicine, University of Innsbruck, Innrain 80-82, 6020 Innsbruck, Austria
* Correspondence: m.rainer@uibk.ac.at

Abstract: Phyto products are widely used in natural products, such as medicines, cosmetics or as so-called "superfoods". However, the exact metabolite composition of these products is still unknown, due to the time-consuming process of metabolite identification. Non-target screening by LC-HRMS/MS could be a technique to overcome these problems with its capacity to identify compounds based on their retention time, accurate mass and fragmentation pattern. In particular, the use of computational tools, such as deconvolution algorithms, retention time prediction, in silico fragmentation and sophisticated search algorithms, for comparison of spectra similarity with mass spectral databases facilitate researchers to conduct a more exhaustive profiling of metabolic contents. This review aims to provide an overview of various techniques and tools for non-target screening of phyto samples using LC-HRMS/MS.

Keywords: non-target screening; LC-HRMS/MS; metabolites; annotation; mass spectral databases

1. Introduction

Natural products represent a big market, with 166 billion US-dollars of sales volume throughout the year 2019 in the US [1], which implies a growth of sales of 4.8% compared to 2018 [1]. This is the reason why phyto-analysis is an important field to assess metabolites in phyto samples for their possible compounds with health benefits and to control their quality in terms of contamination due to their environment, cultivation, or additives, e.g., pesticides, fertilizer, or other toxic compounds. A possible workflow for the detection of metabolites and contaminants in phyto samples using non-target screening is shown in Figure 1. Examples for the need of quality control of phyto samples are pesticides in wine [2] and the addition of diethylene glycol in wine in 1984 in Austria [3], which was used to simulate higher quality and to obtain a higher volume of product [3]. However, due to the Matthew effect, the detection of those contaminants is hampered [4]. This effect describes the psychological effect that decisions on targets of studies are based on occurrence in prior studies, rather than considering that additional factors may have an influence on the studied object [4]. Those compounds represent the so-called "known knowns" and "known unknowns" described with the "Rumsfeld Quadrants" (Figure 2) by Stein [5].

Despite that, real new knowledge is generated by finding unexpected compounds in a sample, which are described as "unknown knowns" and "unknown unknowns" [5]. To find those unknowns, non-target screening represents the method of choice. For this purpose, three major techniques are used to acquire spectra of samples after sample preparation: one of these techniques is nuclear magnetic resonance (NMR) [6,7], the other two techniques represent liquid chromatography coupled to a high-resolution mass spectrometer (LC-HRMS), and gas chromatography coupled to a mass spectrometer (GC-MS) [8]. The latter was used for a long period of time as gold standard to detect volatile compounds,

due to its instrument standardization and, therefore, its comparability of data acquired by different laboratories [9]. A major disadvantage of GC-MS is that polar and thermally labile compounds must be derivatized in order to analyze them [9]. In contrast, LC-HRMS/MS can be used for those compounds with minor sample preparation. In particular, the ability to obtain fragment spectra by collision-induced dissociation, after the ionization process in LC-HRMS [10], which represents a fingerprint of the compound's structure, facilitates the identification. After the acquisition of spectra, data processing has to be conducted to separate compound spectra from each other, using retention times and precursor masses. To be able to conduct non-target screening, spectral databases in combination with search algorithms are needed to compare the acquired fragment spectra with spectra of reference standards for the identification of compounds. However, the occurrence of these reference spectra is a bottleneck. For a total amount of 129 million registered compounds [11], the biggest database for non-volatile compounds includes just more than 40,000 spectra of more than 15,000 compounds [12]. This exemplifies that only a fraction of compounds in samples can be found by comparing LC-HRMS/MS fragment spectra with databases, and more effort must be taken to expand the number of compounds in these libraries. This technique of identification was widely used for the identification of water pollutants [13–17] and drugs [9,18–24]. Jorge et al. published a review describing several targeted applications of LC-MS for metabolite screening [25]. However, non-target screening is a promising technique to conduct an exhaustive metabolite profiling for various applications, such as determining contaminants, biomarkers and the location of cultivation, which will be described in this review. Furthermore, it will outline the different steps of non-target screening, i.e., extraction of the sample, sample preparation, data acquisition, data processing and compound annotation.
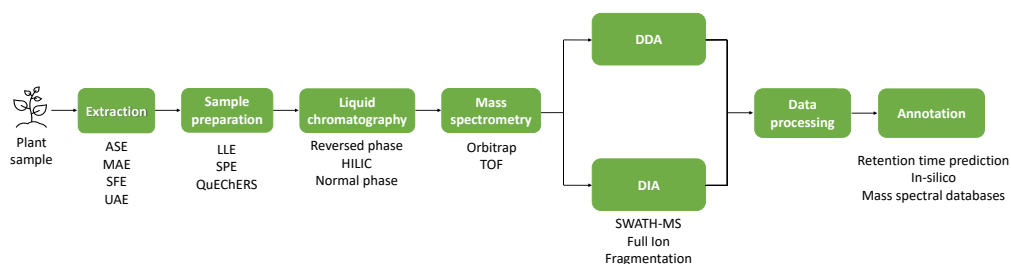

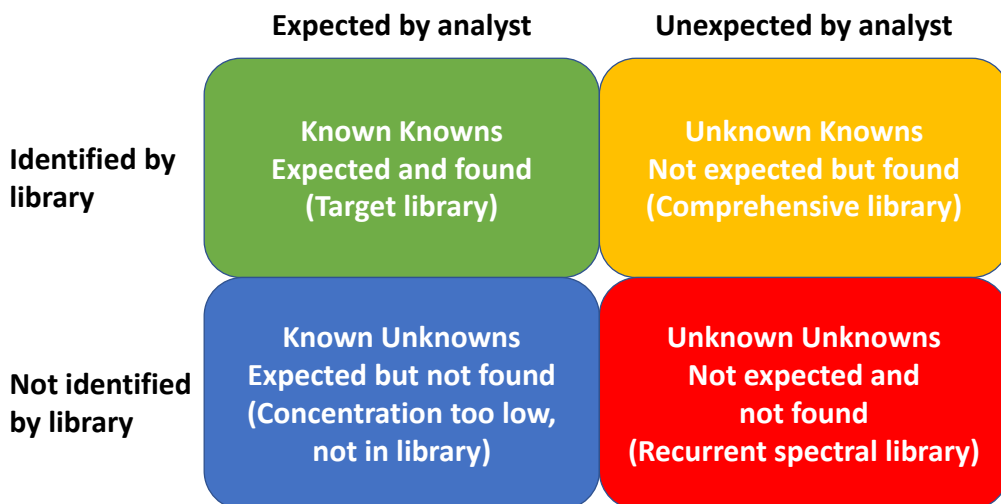**Figure 1.** Workflow for non-target screening.

Figure 2. "Rumsfeld Quadrants" showing the intersection of yes/no answers for whether analysts expect a compound to be identified in the sample (prior probability) and whether it was identified in a library search [5]. Adapted with permission from [5].

2. Sample Preparation

2.1. Extraction of Solid Samples

For the analysis of phyto samples, sample preparation represents the most crucial step. For the extraction of plant metabolites in particular, the first task is to dry or freeze the collected samples to be able to grind them [26]. The smaller the sample particles, the higher the efficiency of the metabolite extraction, due to the increased surface area of the sample. There are several different extraction methods for solid samples, e.g, plants, which will be briefly explained in this section. For a more detailed view on this matter, Azwanida [26] published a review on this subject.

The easiest approach to phyto compound extraction is maceration and infusion. During the process of maceration, the sample is placed into a flask of solvent for several days and is agitated frequently. This leads to a soft extraction of the metabolites, but with a low efficiency, which is why this technique is mostly used to preserve thermolabile compounds. Infusion has the same ground principle; the only difference is that the solvent is heated during the extraction process.

Another efficient extraction technique for phyto samples is the Soxhlet extraction [26]. This method is much faster and needs less solvent for the extraction procedure. The ground sample is placed into the thimble and the solvent in the reservoir is heated. The vaporized solvent condenses in the cooler, which drops on the thimble and extracts the metabolites. If the condensed solvent reaches the level of the siphon, it flows back into the reservoir. This process leads to an exhaustive extraction, although heat sensitive compounds might be degraded.

An even quicker extraction method is represented by the microwave-assisted extraction (MAE) [26,27]. This method needs to be employed with polar solvents, because the radiation only interacts with the dipoles of polar molecules. In addition, radiation destroys the cells of the sample, which eases the extraction of plant metabolites. An advantage of this method is that it can be performed in small centrifuge tubes with a small amount of sample and solvent. However, one must take care to choose proper conditions to avoid the degradation of the compounds [26].

Ultrasound-assisted extraction (UAE) is often used for the extraction of plant metabolites [26,28]. The cavitation of ultrasound leads to an increased contact surface between the solvent and the sample. Furthermore, the cavitation increases the permeability of the cells, which increases the efficiency of the extraction.

All previously described extraction methods are conducted under atmospheric pressure. Additional extraction methods, which use elevated pressure, include pressurized liquid extraction (PLE) and supercritical fluid extraction (SFE) [26,29].

In PLE, the solvent is pumped with an applied pressure from 35 to 200 bar through an extraction cell, which is located in an oven. This oven heats the sample and the solvent up to 200 °C. The augmented temperature and pressure cause a higher extraction yield of the metabolites by the solvent and, therefore, results in a high efficiency. However, the high temperatures might also result in a potential degradation of thermolabile compounds [30].

To overcome this constraint, SFE can be used. Supercritical CO_2 is mostly used as solvent of choice for SFE. For the extraction of thermolabile compounds of phyto samples, a temperature of 31.1 °C and a pressure of 74 bar can be used to transform CO_2 from its gaseous state to a supercritical state, where it simultaneously behaves like a liquid and gas. This state allows the supercritical CO_2 to penetrate the cells of phyto samples and dissolve their metabolites. An advantage of SFE is that the supercritical CO_2 changes its state while depressurizing the apparatus after the extraction, which avoids the laborious step of solvent removal after the completion of extraction [31].

2.2. Extraction and Concentration of Liquid Samples

In the previous section of this article, different extraction techniques for solid samples, such as plant leaves, were described. The resulting extracts tend to have low concentrations of analytes, which can hamper their detection. Therefore, a cleaning and concentration

step is often used, to be able to eliminate interferences, or to separate a certain class of analytes of the extract. The most popular methods for this task will be briefly described in this chapter. A more extensive description of those methods can be found in the reviews of Kole et al. [32] and Bylda et al. [33].

The most common extraction technique for liquid samples represents solid phase extraction (SPE) [27,29,32–34]. This method is fast and easy; it consists of a cartridge filled with a stationary phase, which has to be chosen according to the analyte's properties for their isolation during the separation process. Several specific SPE materials have been developed for different analytes, one of which is Oasis HLB from Waters. Another kind of SPE material are molecular imprinted polymers (MIPs) [29,32]. MIPs consist of a highly cross-linked rigid solid phase or a low cross-linked hydrogel, and both have recognition sites specially synthesized to fit the target compound [35]. This method reveals a high recovery, but is only suitable if a specific analyte has to be extracted.

Liquid–liquid extraction (LLE) is an extraction method which uses the specific affinity for a solvent of compounds depending on their octanol/water partition coefficient (log P) [32,33,36]. Because of that, a mixture of two solvents, e.g., methanol and chloroform, is used to separate the analytes with the two solvents and to clean the extract from impurities. The efficiency of separation can be increased by repeating the process several times, according to the Nernst distribution law [36]. Derived from this method is the salting-out assisted LLE (SALLE) [32]. In SALLE, a salt, for example, $MgSO_4$, is added to a mixture of water and a water-miscible solvent, such as MeOH, which leads to a separation of these two solvents. This technique can be used to separate drugs from a blood sample with ACN.

A combination of SALLE and SPE is QuEChERS (quick, easy, cheap, effective, rugged and safe) [32]. First, the analyte is separated by a SALLE method, using an organic solvent and a salt, then the supernatant is transferred into another vial, in which the sorbent is added to extract impurities.

These are the most common extraction techniques; other methods can be found in the references.

3. Data Acquisition

Besides sample preparation, data acquisition has a huge influence on the ability to identify compounds of a mixture of analytes. To achieve this, liquid chromatography coupled to mass spectrometry (LC-MS) has proven to be a powerful tool to annotate and identify unknown metabolites or other compounds of a sample. Cajka et al. [36], Dudzik et al. [37] and Oberacher et al. [9] reviewed different aspects of data acquisition which can be used for further information. In the following chapter, a short overview is given.

3.1. Liquid Chromatography

Liquid chromatography (LC) represents the first step of data acquisition. It is used to separate compounds, extracted by sample preparation. This is usually achieved by performing LC as high-performance liquid chromatography (HPLC) or ultra-high-performance liquid chromatography (UHPLC) [9]. The difference between these two methods is the particle size and the applied pressure. In HPLC, the particle size is between 3 and 5 μm and a pressure of up to 400 bar is applied [9,36], whereas in UHPLC, the particle size is below 2 μm and an applied pressure of up to 1200 bar is used [9,36]. To separate compounds from each, a column is used. The most commonly used column is a C-18 reversed-phase column (RP). Due to its hydrophobic capacities, it is used to separate non-polar compounds to make the analysis by mass spectrometry easier [36]. Another popular column is represented by the hydrophilic interaction chromatography (HILIC) column. It is used for the separation of polar compounds, and it can also be used for the separation of lipids. Its main advantage is that it can be applied in a broader pH range. Unlike RP, which is mostly used under acidic conditions, HILIC can be used in a pH range between 2.8 and 9 [36]. The mobile phases used for these methods are polar and non-volatile, which is an advantage for the

coupling to MS, unlike normal phase chromatography (NPLC) [36]. A combination of the latter mentioned techniques, RP and HILIC, is represented by two-dimensional (2D)-LC. It can be applied in an online or offline mode. Two-dimensional-LC in online mode can be automated; however, this results in a lower resolution of the second dimension, due to a shorter separation time [36]. Another important fact to mention is that a powerful software tool for peak alignment is needed, because in offline mode, a series of 1D chromatograms are obtained. For this case, a 2D feature detection algorithm is implemented in MZmine 2 [36]. Another promising separation technique for lipid and metabolite profiling is supercritical fluid chromatography (SFC) [36], but it requires complicated apparatuses and high costs. The measured retention times of analytes can be used to identify compounds, which will be described in the upcoming chapters.

3.2. Influences on Ionization

Ionization is the next step of data acquisition. Due to a high efficiency in ionizing organic compounds, an easy coupling to LC and, therefore, a possible high throughput, electrospray ionization (ESI) is usually used [38]. Another popular ionization method is atmospheric pressure chemical ionization (APCI) [38,39]. In the process of ionization, many variables influence the quality of the respective ionization technique. Matrix effects represent the biggest influence on ionization [39,40], and lead to either an ion enhancement or an ion suppression. For a more detailed view on matrix effects, Niessen et al. [39] and Peters et al. [40] reviewed this aspect. APCI ionization has shown to be less prone to matrix effects [40], but because of its higher sensitivity, ESI is used more often. With an appropriate sample preparation, and the choice of a suitable mobile and stationary phases, matrix effects can be minimized. However, one must keep in mind that they are always present during analysis. To assess the extent of matrix effects, the method of choice is to use a stable isotope-labelled internal standard of the compound, which should be analyzed [39,40]. These standards co-elute with the analyte and the relative amount of lost internal standard can be considered to be the same as the loss of analyte. However, the use of those standards is not always applicable due to the limited presence of suitable standards and their high costs [39,40]. Another method of taking matrix effects into account without the need of using isotopically labelled standards is by adding a standard to be used as the calibration method, but this method is time consuming and laborious, because a new set of calibration solutions must be prepared for each sample [40]. The evaluation of the matrix is important to obtain information about the sensitivity of the acquisition method. If the sensitivity of the method is poor, it could hamper the annotation of unknown compounds by the algorithms and would, therefore, not be suitable for non-target screening [39,40].

Besides matrix effects, the size of droplets produced by the ionization source also proved to influence the ionization. Bahr et al. [41] showed that ESI sources are prone to more ion suppression than nano-ESI sources. They described that, due to a bigger size of solvent droplets in an ESI source compared to a nano-ESI source, the surface area which can be used for compound ionization is smaller and, therefore, less compounds are ionized. This mechanism leads to a higher suppression.

Liigand et al. described the link between the pH value and the ionization efficiency of analytes [42]. In their research, they analysed 28 analytes with 22 different solvent mixtures at pH values of 2.1 and 7.0. They concluded that analytes which tend to be pH independent during the ionization process are either very hydrophilic, whereby the analyte is mostly in the center of the droplet, or highly hydrophobic, where the analyte can be found mostly at the surface of the droplet. Compound polarity, the ability to accept hydrogen bonds and the number of charge centers are the reasons why ionization efficiencies are influenced by solvent pH values [42]. In a study performed by Liigand et al., aniline revealed the biggest difference in ionization efficiencies.

Kruve et al. [43] described the dependency of adducts and additives in the mobile phase. The authors analyzed 17 mobile phases with 9 different adducts each, and measured the adducts which occurred. They found that oxalic acid, as an additive, leads to more H^+

adducts, whereas acetic acid and formic acid tend to promote the formation of Na^+ adducts. The authors encourage the community to use additives in their measurements, due to increased repeatability. By adding additives, the formation of adducts can be controlled better and possible variations can be avoided [43].

3.3. Mass Spectral Analysis

In performing a non-target screening, high-resolution mass spectrometry (HRMS) has proved to be a powerful tool for identifying unknown metabolites in complex matrices [9,36]. These spectrometers have a resolving power between 10,000 and 450,000 full width at half maximum, and a mass accuracy of under 1 to 5 ppm [36,44]. A more detailed review of these apparatuses was published by Cajka et al. [36] and Oberacher et al. [9]. For non-target screening, the MS/MS technique is usually used for molecular fingerprinting due to the fact that each compound is fragmented in a specific way, which results in a unique fragmentation pattern [9]. Fragmentation is performed usually by collision-induced dissociation (CID) in quadrupole time-of-flight (QTOF) instruments or by higher energy collisional dissociation (HCD) in Orbitrap instruments. In the next step, the fragmentation spectra can be compared with published data or with data published in publicly available mass spectral libraries, such as MassBank [45] or Metlin [46,47]. The workflow of spectral annotation will be discussed in Section 5.5.

The two most common high-resolution (HR) MS instrument types for non-target screening are QTOF and Orbitrap analyzers [9]. Oberacher et al. [48] showed that data acquired by one of these instrument types can be used to annotate compound spectra acquired by the other instrument type. First, the authors compared with each other the data of their own mass spectral databases, WRTMD [49] and Eawag [45], and concluded that data acquired with a collision energy between 20 and 50 eV on the QTOF instrument and a nominal collision energy of 30 to 60 on the Orbitrap instrument provided the best matching results for the overlapping of compounds of both databases [48]. Furthermore, these results were evaluated with effluent samples of wastewater treatment plants, which proved this conclusion.

All data of the mass spectral databases mentioned in the latter were acquired in the so-called data-dependent acquisition (DDA) mode (Figure 3) [44,48]. DDA is mostly used for non-target screening. In the DDA mode, a full scan is first acquired over the entire mass range, then the ions of the most intense signals above a chosen threshold of the full scan are submitted to the fragmentation cell. The biggest disadvantage of this method is that, due to the selection process of ions, only a fraction of all possible compounds is fragmented and can, therefore, be found in the screening process [9]. As explained in the previous chapter, pH and the choice of solvent can influence the ionization efficiency, which can hamper the detection of ions because of a possible suppression of an analyte and a resulting drop in intensity below the chosen threshold, which causes the prevention of the fragmentation of the analyte [36]. The reason why DDA is usually used for non-target screening is its ability of a high throughput of samples and its high sensitivity [9]. In addition, the precursor mass of the respective ion remains directly linked to its fragment spectrum, which makes the annotation much easier.

In contrast, data-independent acquisition (DIA) methods fragment all ions of the sample (Figure 4). The simplest mode is the all-ion fragmentation technique, which fragments every ion that is ionized. The sequential windowed acquisition of all theoretical fragment ion mass spectra (SWATH-MS) represents another DIA method, and was described by Gillet et al. (Figure 5) [50]. In SWATH-MS, all ions in a certain mass window are fragmented. The mass window described by Gillet et al. was set to an m/z of 25 Da plus a mass overlap of 1 Da for each window [50]. The authors compared the method for the analysis of yeast proteins with single reaction monitoring (SRM). It was concluded that targeted identification of the proteins is comparable with SRM analysis [50]. The biggest disadvantage of SWATH is the deconvolution process which is needed to separate the fragmentation spectra of the different ions from each other [36]. This can be achieved by

employing software tools such as MS-DIAL of Tsugawa et al. [51]. Another disadvantage is that just a limited number of mass windows per chromatographic peak can be acquired, because of the amount of time needed for the collection of an MS1 spectrum (up to 100 ms) and the corresponding collection of a SWATH spectrum (10 ms per window). Therefore, only a maximum of 40 windows can be acquired per chromatographic peak [51], which limits either the mass range, or the range of a mass window. In addition, is still a limited number of instruments capable of performing SWATH acquisition [52].
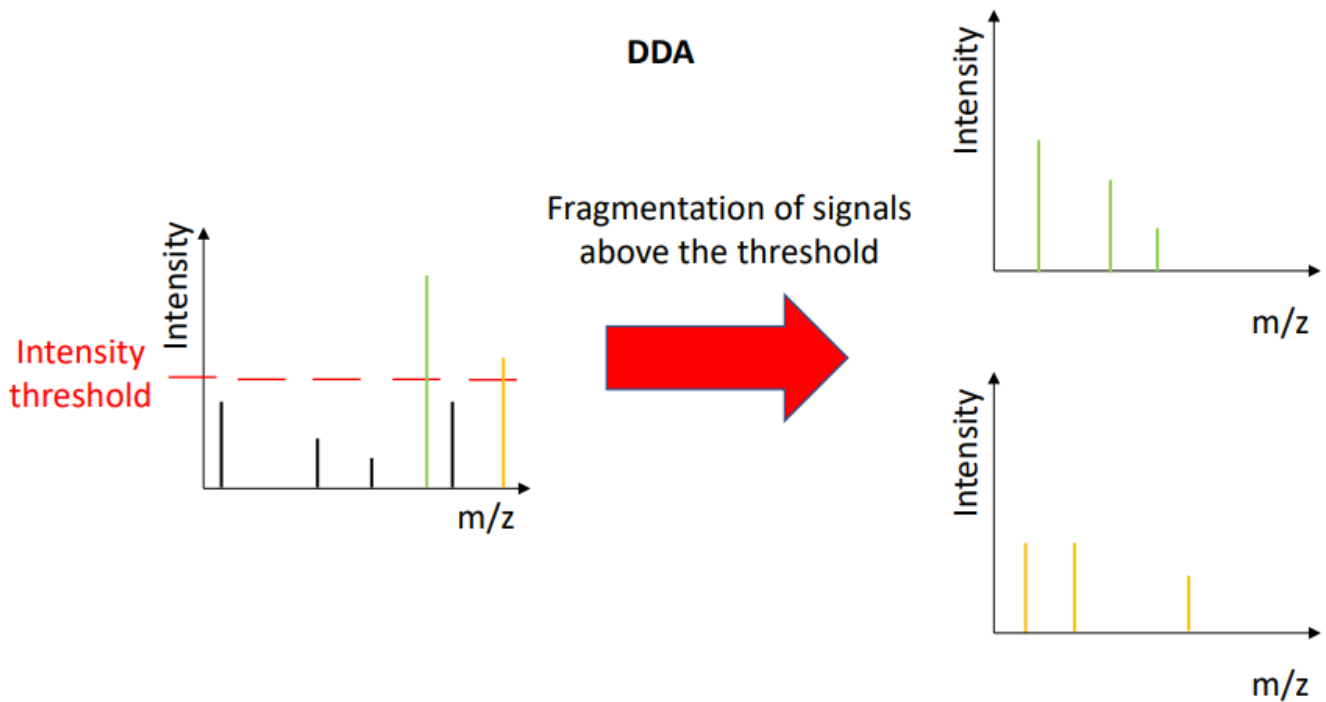

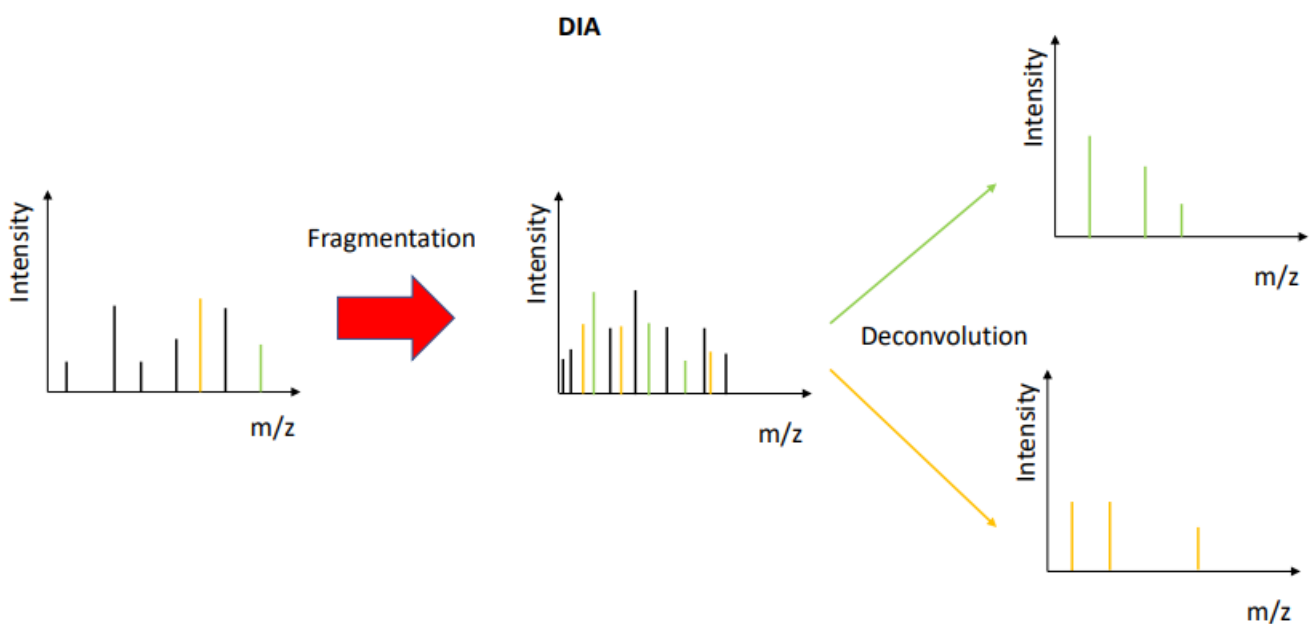**Figure 3.** Schematic of DDA fragmentation of two compounds (green and yellow).

Figure 4. Schematic of DIA fragmentation of two compounds (green and yellow).
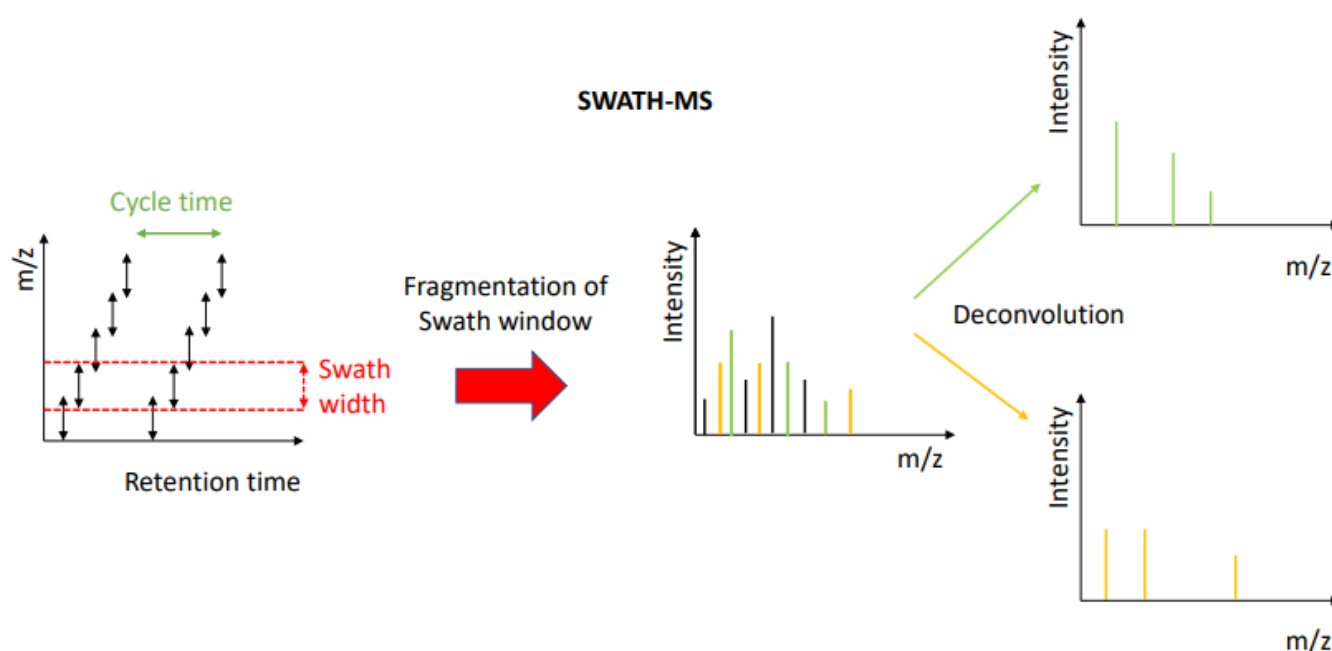
Figure 5. Schematic of SWATH-MS fragmentation of two compounds (green and yellow).

Guo et al. performed a comparison of acquisition modes using full-scan, DDA and DIA in all-ion fragmentation and SWATH modes by employing the different modes on a urine sample [52]. The authors assessed the different modes with five different criteria: number of features, convenience, quantitative precision, MS/MS spectral quality and MS/MS spectral coverage. In terms of the number of features, full-scan mode performed best, which is not surprising, because all of the acquisition time is used for feature elucidation and, therefore, most of the features are detected. Hence, the use of full-scan mode leads to limited information in untargeted screening, which makes annotation of unknowns exceedingly difficult, because only retention time and precursor mass can be used for this task. According to the authors, the quantification precision with the full scan was as good as it was with DIA methods [52]. However, due to additional structural information, the authors stated that DIA should be used for targeted quantitative analysis. The previously stated features can be assessed by using MS without fragmentation. Considering MS/MS experiments, Guo et al. stated that DDA leads to fragmentation spectra having a higher quality, which supports the annotation process and explains why DDA is the method of choice for non-target screening, although DIA showed higher MS/MS spectral coverage. According to the authors, DDA proved to be the best method in terms of convenience, because there are more tools available for non-target screening of metabolites and, therefore, the annotation process can be better automated [52].

An additional DIA approach, other than SWATH, was described by Broeckling et al. [53]. This method is called "indiscriminate MS/MS" or idMS/MS. This method uses the differences between spectra acquired with high and low collision energies, whereas spectra collected with low collision energies correspond to the unfragmented spectrum in the full-scan mode. This is the reason why the disadvantage of the loss of correlation between the precursor ion mass and the respective fragment spectrum can be overcome with the use of computational tools. In the article, the authors used XCMS [54] and CAMERA [55] to find correlations between the full scan and the fragment spectra, by aligning them with features of their retention times and peak shapes. Additionally, the authors employed their workflow on spiked barley extracts and on 183 blood samples. In the assessment of the spiked barley extract, the authors stated that the identification of the spiked caffeine standard was not successful due to an inaccurate integration of its peak, which led to a false retention time and, consequently, influenced the feature detection of the software tools [53]. In comparison, the reconstructed spectra collected from

183 blood samples, which were annotated with MassBank showed a similarity of 0.911 for the caffeine spectrum. This workflow could be used for a more exhaustive metabolite screening, because all ions in the acquired mass range are fragmented. However, due to the sensitivity problems shown in the assessment of the barley extract, this method should only be used in addition to DDA when performing non-target screening.

Broeckling et al. also described another acquisition method based on DDA called data-set-dependent MS/MS (DsDA) [56]. This workflow employed a classical DDA in the first step. After the measurement, a feature detection step was performed with XCMS. Following this step, a score algorithm was developed by the authors, which gives a higher score to features which possessed high quality features in full-scan mode, but low-quality features in MS/MS [56]. Features which received the highest scores were those with the highest difference in feature quality between full scan and MS/MS spectra. These features with higher scores were prioritized in the next injection for fragmentation. In addition, a MaxDepth [56] option was added which forced those features which had not been fragmented so far to be fragmented every fifth scan. This approach was tested on 33 complex samples and 7 QC samples. In this experiment, the recognition of reproducible features of these 40 injections was augmented from 47% with DDA to 72% with DsDA, and up to 80% with MaxDepth [56]. This shows that different acquisition techniques, in combination with novel data processing tools, can produce further improvements for feature recognition and, consequently, for annotations in non-target MS screening.

In the next chapter, different software tools for data processing will investigated in more detail.

4. Data Processing

After obtaining recorded mass spectra with the previously described techniques, raw data has to be processed with appropriate tools to be able to annotate analytes with non-target screening [44]. Katajamaa et al. described the theoretical basics of data processing in detail in their review [57]. Briefly, the steps of data processing are filtering, feature detection, alignment and normalization [57]. The filtering step includes noise reduction by applying smoothing functions, such as the Savitzky–Golay filter. In addition, baseline removal can be applied to remove the shift of the baseline caused by chemical noise [57]. Feature detection is the next step in data processing. There are three different approaches for feature detection. The first approach is vectorized peak detection, where peaks are independently found in the mass and retention time direction [57]. The next approach is extracting data to produce an extracted ion chromatogram (XIC) with a narrow mass window. These XICs are independently searched in the time domain for peaks, which represent features [57,58]. The last approach for feature detection is fitting a model against the raw data. Here, isotopic patterns are included to find features, which could improve the feature detection [57]. Alignment is needed for the correction of retention times between different runs of a sample or between recordings of different samples. Alignment algorithms use either all recorded raw data or cluster features in a matrix, where the rows represent features and the columns represent different additional information, such as peak area [57]. The aim of normalization is to remove bias in ion intensities caused by systematic differences between measurements of the samples. Two major approaches are applied for normalization: the first is the statistical approach, which aims to find a statistical model for scaling of intensities. The second approach is empirical, by normalizing the measurements with signals recorded from internal or external standards [57].

The above is a short overview of data processing techniques. In the following section, some algorithms for data processing are introduced and several articles on the comparison of those software tools will be described.

A major part of the XCMS [54] package is the "centWave" algorithm for feature detection described by Tautenhahn et al. [59]. The algorithm has three input parameters: mass deviation, in parts per million (ppm); chromatographic peak width range, in seconds (s); and the signal-to-noise threshold. At the beginning, the algorithm detects certain regions

of interest (ROI) in the m/z domain. Then the base and noise lines are estimated. Next, chromatographic peaks are detected by applying continuous wavelet transform (CWT), which has an advantage in detecting peaks of widely differing widths [59]. Furthermore, the boundaries of the chromatographic peaks are located and the feature intensity within these boundaries is calculated. As a last step of the algorithm, the centroids of the m/z features and the signal-to-noise ratio are computed [59]. The authors compared this algorithm by applying "matchedFilter" of the XCMS package [54] and centroidPicker of MZmine [60] on a dilution series of extracts of *Arabidopsis thaliana* and a mixture of 14 standards. First, parameter optimization was conducted to obtain the best results for each of the algorithms. With these results, an F-score (a combined measure of recall and precision) was calculated for each of the three algorithms [59]. The results showed that, overall, the centWave algorithm demonstrated the best F-score for all samples and needed less time for data processing than both other algorithms [59].

Recently, McLean et al. published an automated tool "AutoTuner" for parameter optimization for data processing software [61]. Parameters to be optimized by this algorithm are group difference, ppm, signal-to-noise threshold, scan count, noise, prefilter intensity and peak width for the centWave data processing algorithm [61]. An extensive description of the optimization process can be found in the article [61]. The authors compared their optimization procedure with the most commonly used optimization tool, "isotopologue parameter optimization" (IPO), by applying both of the algorithms on a set of 85 metabolites. IPO optimizes the parameters by running the centWave algorithm iteratively until the scoring function of the algorithm is maximized. The benchmarking of 85 standard metabolites showed a better performance in terms of accuracy for the AutoTuner algorithm, compared to IPO. AutoTuner detected 100 out of 101 possible features of the sample, whereas IPO detected 82 features of this sample of standards. In the next step, the authors compared the performance of these algorithms by applying them on a bacterial culture and a rat fecal microbiome [61]. In positive mode, AutoTuner detected 203 unique features compared to 2606 unique features detected with IPO. A similar result was observed in negative mode, where 540 unique features were detected with AutoTuner and 3420 unique features with IPO [61]. This is the reason why the authors were able to find more features in MS/MS libraries, which is not surprising. In spite of that, the biggest advantage of the AutoTuner optimization algorithm is its speed for calculating the parameters. For example, the authors just needed 25 min for parameter optimization of the fecal microbiome samples with AutoTuner, whereas IPO needed more than 38 h [61]. This shows that the algorithm could be useful if quick feature detection is needed, but in terms of accuracy, IPO is powerful, especially with complex matrices. However, this approach could be a hint that more sophisticated algorithms can improve the calculation times of data processing and, at the same time, improve the quality of this process. To our knowledge, a comparison using these parameter optimization algorithms with plant samples has not yet been published.

Due to the vast amount of data processing, from a researcher's perspective, it is important to know which data processing software tools perform best. Thus, some articles are presented which compare vendor and open-source data processing tools with each other.

Rafiei et al. compared the commercial data processing tools Peakview®, Markerview™ and MetabolitePilot™ from AB Sciex with the open-source tool XCMS Online (Figure 6) [62,63]. A mixture of 84 metabolite standards was recorded by LC-HRMS and processed with the four different software tools. The authors employed similar parameters for the peak-picking algorithms, with a 5-ppm mass accuracy, a 500-counts-per-second (cps) intensity threshold and a minimum peak width of 5 s. The found peaks were processed with Matlab [64] to find overlapping peaks in the produced peak lists and in the next step, they were annotated using the Metlin database [46]. Altogether, only 24 metabolites out of 84 were matched in positive mode applying the four peak-picking algorithms. In negative mode, this number decreased to just 13 metabolites. The best recovery rate was achieved using Markerview, with 23 identified metabolites out of 24, and 13 metabolites present in

all four workflows [63]. In the next step, the workflows were applied on bile and urine samples. XCMS and Markerview found the most peaks in these samples, with 11,708 and 11,553 peaks, respectively. In comparison, Peakview performed poorer, with just 2015 peaks found, but yielded the highest amount of tentatively matched metabolites in Metlin. Markerview also showed the highest two- and three-way overlap of 9.9% and 10.6%, respectively [63]. In addition, the authors stated that XCMS Online and Markerview were faster than the other workflows, and they concluded that Markerview performed best in this study [63].
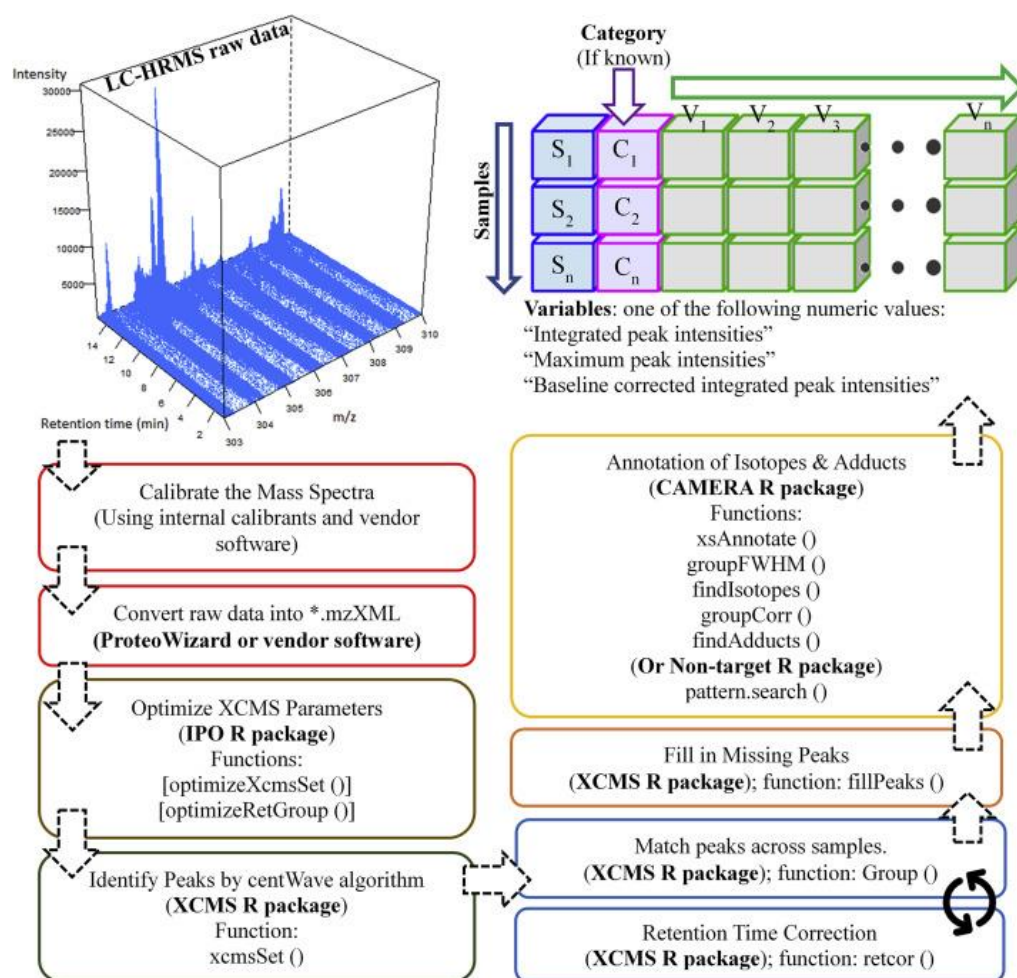

Figure 6. Peak-picking workflow with XCMS [34]. Reprinted from [34] Copyright 2020, with permission from Elsevier.

Recently, Smith et al. compared the performance of the Massifquant [65] algorithm from the open-source software tool MaxQuant [66], the previously described centWave [59] algorithm, the MZmine 2 [67] workflow and the MatchedFilter [54] algorithm, by applying these algorithms on XICs of a set of 48 manually curated human proteins of the universal proteomics standard, organized in six abundances [68]. A brief description of the workflows of these algorithms can be found in the article of Smith et al. [68]. Optimization of parameters for the centWave and MatchedFilter algorithms was conducted by applying IPO. For Massifquant and MZmine 2, the parameters were optimized by iteratively changing them and using the parameters, which proved to give the best results. Tryptic digestions of the samples were recorded in the DIA method with a QTOF instrument. Finally, the authors compared the results of their manually curated XICs with the results of the algorithms. The result of this study shows that only MZmine 2 and Massifquant recognized more than 50% of all features, which were manually found, with a true positive rate of 76.6% for MZmine

2 and 66.7% for Massifquant. In comparison, MaxQuant performed poorly, with a true positive rate of just 1.9% and the highest number of false negative features. In terms of false negative features, MZmine 2 performed best with just 13,963 features [68]. This shows that open-source tools can perform quite well, but one must carefully choose an appropriate software to obtain the best result.

Hohrenk et al. compared the performance of the algorithms of open-source tools MZmine 2 [67], XCMS Online [62] and enviMass [69,70]. In addition, the vendor software "Compound Discoverer" by Thermo Scientific was used for the comparison study [70]. The parameters for each algorithm were optimized by measuring the recovery rates of internal standards and suspects [70]. Samples for this study consisted of surface water samples of the Diemel river and samples of a wastewater treatment plant, and measurements were performed on an Orbitrap instrument. All samples and a blank were measured in triplicate. Several filtering steps were conducted, such as blank subtraction, componentization and a replicate filter. The authors stated that these filtering steps have a big influence on the outcome of the peak-picking process and produced divergences [70]. As a result of this study, 10% of all features were identified by all algorithms, whereas 40 to 55% were found by only one of these algorithms. The authors concluded that for every non-target screening, the appropriate peak-picking tool has to be chosen, as each algorithm has its advantages and disadvantages [70].

Another tool for correct annotation is the correct assignment of molecular formulas. Kind et al. published the seven golden rules for a potential correct assignment of molecular formulas [71]. The most important prerequisite for this task is to record HRMS data. These rules will be briefly described. The first rule is the restriction for element numbers. This rule implies that for a certain accurate mass, only a certain number of atoms of an element are possible. Furthermore, the second rule is the Lewis and Senior check. The Lewis check implies that all σ and p shells are completely filled. The Senior check has three conditions: the number of atoms with odd valences is even, the sum of valences should be greater or twice the number of maximum valences and, finally, the sum of valences is higher or twice the number of valences minus 1 [71]. The third rule is the inclusion of the isotopic pattern of the atoms, because halogenic atoms can, in particular, be determined by their specific isotopic pattern [71]. In addition, the ratio of an atom to the number of C-atoms can be used for the determination of the molecular formula, which represents rule four. For example, the H/C ratio is usually between 3.1 and 0.2 [71], which covers 99.7% of all compounds. The authors included a table with other common ranges for the most common atoms [71], which explains the fifth rule, the heteroatom ratio check. This kind of check aims for the exclusion of compound formulas which have too high a number of heteroatoms, e.g., the common range of N/C is 1.3 to 0. Despite rule five, several distinct formulas would be possible, which is why the authors included a threshold to limit the combination of the heteroatoms N, O, P and S which represents rule six [71]. Finally, rule seven is the check for trimethylsilyl (TMS) groups, because these groups are often used for capping acidic groups and are, therefore, replaced in the algorithm of the authors by an H atom [71]. The algorithm was checked by randomly choosing spectra of databases and applying the algorithm on them, which resulted in 98–99% of correctly assigned formulas by querying the correct database, and 84–90% correct formulas by querying PubChem [71]. Kind et al. implemented the algorithm as a script in Microsoft Excel [71].

A more recent view on the techniques of molecular formula assignment can be found in the review of Ljoncheva et al. [72].

5. Non-Target Compound Annotation

In the previous chapters, the workflows of sample preparation, data acquisition and data processing were described, which only generates a huge amount of raw data. However, the aim of non-target screening is to annotate compounds in the sample, which proved to be difficult due to the vast number of possible compounds in the chemical space. This chapter gives an overview of recent developments in annotation workflows.

5.1. Compound Identifier

Compound identifiers are crucial for the annotation of unknown compounds. The most commonly used identifier for compounds is retention time, which is recorded by the analyzer of HPLC or UHPLC instruments, and is unique to the combination of the instrument settings, the stationary phase and the used mobile phase. Another identifier is accurate mass, which can be accurately determined by HRMS. This accurate mass can be used by the previously described algorithm to propose molecular formulas of the unknown compounds to generate a suspect list for these compounds and make the annotation process easier. If a suspect list of compounds is available, the search for spectra in libraries can be facilitated by searching with InChI or InChIKey [73,74]. InChI is the older standard of compound representation introduced by IUPAC. It consists of the following five layers: formula, connectivity, isotopes, stereochemistry and tautomers [73]. This representation can also be interpreted by the user. If one is familiar with the structure of this representation, the molecular structure of the compound can be read. For standardization matters, InChI should always begin with the letters "InChI = 1 S/", where the "S" indicates that the code that follows is in standard form [74]. In contrast, InChIKey is not directly readable by the user, because it is calculated using the corresponding InChI with a SHA-256 hash [73] and should facilitate internet and database searches. It consists of three blocks separated by two dashes having a size of twenty-seven characters [74]. The first block of fourteen characters encodes the skeleton of the compound structure, the first eight characters of the second block encode additional information of the structure, such as the stereochemistry or isotopic substitution [74]. If the ninth character of the second block is an "S", this means that the corresponding InChI is in standard form, otherwise, this character would be a "N" [74]. The tenth character of the second block indicates the version number of the InChIKey algorithm, where an "A", for example, means version 1 [74]. The third block consists of just one character, which represents the protonation of the compound, whereas "N" means neutral, and "O" indicates a +1 protonated compound [74]. These identifiers should be included in the workflows to ease the indication of the exact compound structure for scientists and readers. A good website for the generation of InChIKeys is Chemspider [75], where the user can translate InChI to InChIKey. Another website for this task is the EPA CompTox dashboard, where a batch of compound names can be translated into the corresponding InChIKey [76].

5.2. Confidence Levels of Annotation

Schymanski et al. proposed five levels for communicating annotation confidence (Figure 7) [77]. Level 1 represents a confirmed structure, which was confirmed with reference standards by recording its MS, MS/MS spectra and retention time. For non-target screening of metabolites in complex samples, level 2, at least, should be the goal in terms of its annotation confidence, which represents a probable structure. This level is divided into two sublevels. Level 2a needs an unambiguous match in a spectral library and a match of retention time or retention time index, which is often used in gas chromatography [77]. Level 2b needs confirmation with experimental data. This level can be obtained if no spectral data is available in the literature or spectral library, but the acquired experimental data excludes any other structure. Tentative structures represent a confidence of level 3. This level is similar to level 2b, but the exact structure cannot distinguish which possible isomer is the right one. Another example of level 3 identification would be high ranked matches of in silico fragmentation. Unequivocal molecular formula represents confidence level 4 [77]. This is the case where a molecular formula can be determined using an MS spectrum, but, for example, there are too many impurities in the spectrum to be able to use the spectrum for a more detailed annotation. A level 5 confidence is reached by comparing the accurate mass with the measured data [77].

Identification confidence		Minimum data requirements
Level 1:	**Confirmed structure** By reference standard	MS, MS², RT, Reference standard
Level 2:	**Probable structure** a) By library spectrum match b) By diagnostic evidence	MS, MS², Library MS² MS, MS², Experimental data
Level 3:	Tentative candidate Structure, substituent, class	MS, MS², Experimental data
Level 4:	Unequivocal molecular formula	MS isotope/adduct
Level 5:	Exact mass of interest	MS

Figure 7. Proposed identification confidence levels in high resolution mass spectrometric analysis. Note: MS² is intended to also represent any form of MS fragmentation (e.g., MSe, MSn) [77]. Adapted with permission from [77] Copyright 2014 American Chemical Society.

5.3. Retention Time Prediction

Retention times can be used to rule out false structures matching to the recorded accurate mass. This technique is used in gas chromatography (GC) by applying retention time indices (RI), which, in this case, can be achieved because of the standardization of instrument settings of GC [72]. In comparison, many different settings are used in LC-MS, which makes this approach almost impossible to apply. However, retention times can be used to differentiate between different isomers which have the same accurate mass [58]. For this task, different approaches were developed, such as quantitative structure−retention relationship (QSRR) models, artificial neural networks (ANN), a deep-learning regression model (DLM) and a mapping approach between different chromatographic systems, which will be briefly described in the following section.

QSRR models predict retention times by comparing experimental retention times of reference compounds and their physicochemical properties with the properties of suspects [58,72]. Creek et al. described a QSRR model for HILIC columns, which helped to remove 40% of falsely identified compounds [78]. The authors used recorded retention times of 120 metabolites at a pH of 3.5, and with multiple linear regression determined that six physicochemical properties had an influence on the compounds' retention times [78]. Out of these six properties, the octanol–water partition coefficient log D was the most influential variable of this model. It is stated in the article that this model should only be applied on metabolites with a mass below 400 Da, because all standards with a mass above this value were excluded due to the poor prediction of log D [78]. Furthermore, to demonstrate the performance of their prediction model, the authors conducted a non-target screening on standard solutions with four different concentrations. The spectra of these solutions were matched against metabolite databases which led to 3133 putative identifications, although the standards consisted of just 127 metabolites. Therefore, an additional filter step was applied, using the QSRR model for retention time prediction. This step ruled out 40% of the original falsely identified compounds. In addition, the model was also applied on cell extracts of *T. brucei*, where it could remove 35% of false positive identified compounds [78].

Recently, Aalizadeh et al. published another QSRR model for RP and HILIC stationary phases [79]. The authors described three different models, one for RP-LC columns coupled to an ESI-source ionizing in positive mode as well as in negative mode, and one model for HILIC-LC columns coupled to an ESI-source ionizing in positive mode [79]. All prediction models were based on different sizes of training and validation sets. The RP-LC

model coupled to the positive ESI-source used a training set of 1461 compounds and was validated with 369 compounds, with five physicochemical variables showing the highest correlation for the retention time prediction. In contrast, the model of RP-LC coupled to an ESI-source operating in negative mode used just 247 compounds as the training set and 62 compounds as the validation set, leading to a model consisting of eight variables. Due to the difference of the size of their training sets, the model for the positive mode ESI showed a better prediction result. More than 70% of the predicted retention times in positive mode were within a 1-min time window around the experimentally determined RT [79]. It was determined that the prediction model of HILIC-LC coupled to an ESI-source operating in positive mode showed the best results with seven physicochemical variables. It was built on a training set of 542 compounds and a test set of 140 compounds. The results show that 71% of the dataset was within a 1-min time window compared to experimental data. In addition, using Monte Carlo sampling, Aalizadeh et al. determined acceptable error windows to be 12% of the overall LC run time [79]. This was accomplished by retrieving experimental data from MassBank [45] and building thirteen different QSRR models. The 95th quantile mean of predictive residuals was calculated with Monte Carlo sampling, which resulted in the acceptable error windows [79]. Furthermore, the performance of the aforementioned models was illustrated by identifying transformation products of Tramadol and Niflumic acid, which have the same mass, but different retention times. By applying RT prediction, these isomers were successfully annotated because of the differences in their RTs [79]. Another application of these models was described, which was the identification of biocides in influent, effluent waste waters and sewage sludge, where the prediction models helped to narrow the list of suspects and, therefore, led to a more precise annotation of compounds [79].

In contrast to the QSRR approach, Stanstrup et al. chose another approach for predicting RTs [80]. The authors developed a tool called "PredRet", where retention times can be shared and compared with other laboratories. This data is used to create a model to predict the correlation of retention times between different chromatographic systems. Such a model is created if at least ten common compounds are reported for two different systems [80]. These models are regularly updated when more data has been submitted to the tool. In the article, the authors stated that the original models were developed using data of 23 different databases of which 2 were recorded on a HILIC-LC system [80] and data of all other systems were recorded on RP-LC systems. To avoid the inclusion of erroneous data in the prediction model, the authors implemented a weighing function, which penalizes residuals above a threshold of 10% [80] in the modelling step. In the prediction step, another function was implemented that detects compounds which have twice the prediction interval width of the predicted retention time value, and uses all the other models to determine if the prediction is wrong. If this compound is included in just one of the models, both of its data points are discarded [80]. The prediction error of this approach for all compounds was below 0.28 min, with a mean error of 2.6% [80].

Hall et al. used artificial neural networks for creating a RT prediction model for RP-LC [81]. The authors used bitkey analysis, where each bitkey represents a feature present in the structure of the compounds. So, the authors chose the compounds of their training set depending on the best coverage of bitkeys in order to predict the RT of complex metabolites with the ANN by using simple metabolites which possess structural combinations of features [81]. Overall, 1955 compounds were used to build this model and 202 human metabolites were used for independent validation. The training set was divided into 278 classes of compounds having the same features and were ordered by their complexity. These classes were used to generate two different models applying ANN [81]. The exact procedure of model building can be found in the article [81]. Prediction models had a true positive rate concerning the independent validation set of 93% and 94%, respectively. This shows that an ANN approach could be promising; however, more work has to be undertaken to cover more compounds of the chemical space.

The previously described models were constructed with data of no more than 2000 compounds. Recently, Domingo-Almenara et al. described a model which used 80,038 small molecules to build a retention time prediction model for RP-LC with DLM [47]. All standards were specially assembled by experts, where 75% of these molecules were randomly selected for the training set and the validation set consisted of the other 25% of molecules [47]. In their article, the authors compared the similarity of molecules with the Tanimoto coefficient [82] and applied the DLM on this approach. The authors stated that a similarity of more than 90% between the compounds in the training and validation sets improved the prediction error significantly [47]. Only one compound with 90% similarity in the training set was needed to observe such an effect. Next, the authors used their model for the evaluation of the prediction of RTs in other chromatographic systems. To achieve this, data from four chromatographic systems of the previously described PredRet tool, which had the largest set of data, were taken. For the task to predict the RTs in another chromatographic system, the authors used 50 randomly chosen molecules for the projection of one system to another [47]. By doing so, the projection yielded a median relative error of 8 to 10% for short chromatographic systems, and 14 to 17% for long systems [47]. In the next step, a threshold for each system was determined, which was used to discard all molecules of the predicted projection. By applying this threshold, approximately 70% of the correct molecules were ranked in the top three [47]. This shows, as described earlier, that RT prediction could help in the annotation process, and that a projection to other systems might facilitate researchers to gather information of compounds, for which the laboratory has no reference standards.

5.4. Annotation with In Silico Fragmentation

For the same reasons, in silico fragmentation is used to cover molecules of the chemical space for which no reference standards are available [83]. There are three major approaches for in silico fragmentation [83–85]. Rule-based methods use physicochemical properties to predict fragmentation patterns, similar to the QSRR approach in retention time prediction, whereas combinatorial methods use substructures of molecules to predict certain fragment peaks by searching for their specific fragmentation trees [84]. A tool which uses this approach is MetFrag [86]. The third approach for in silico fragmentation is machine learning. This approach requires training by a big training set and will be improved with every MS/MS spectrum recorded [84]. CFM-ID is a well-known tool which uses the machine-learning approach [87]. The biggest constraint of the above methods is that they are preferably used for small molecules only, because calculations of bigger molecules need much more time [84]. A more thorough description of these tools and mechanisms can be found in the following references [44,58,72,83–85,88]. In the next section, a few of these tools and their comparison will be described.

Sirius 4 [89] is such a tool, which uses isotope patterns of MS/MS spectra to calculate the respective fragmentation trees using deep neural networks (DNN) [89]. To build the model for this tool, 11,728 compounds were used as the training set. Furthermore, the authors included CSI:FingerID [90] for scoring the predicted fragmentation trees [89]. In addition, users can narrow the search by including a self-assembled suspect list. The biggest advantage of this tool is its quickness of the calculation process. In this article it is stated that, for the calculation of 3965 compounds, Sirius 4 needed less than 6 h [89].

Another previously mentioned tool for in silico fragmentation is CFM-ID 3.0, described by Djoumbou-Feunang et al. [87]. As training sets, experimental data of KEGG [91], MoNA [92], PhytoHub [93], DrugBank [94], HMDB [95] and GNPS [96] were used. These data were used to calculate fragment spectra with three different collision energies (10, 20 and 40 eV) for more than 108,000 compounds. In addition, for version 3.0, the authors developed fragmentation rules for lipids to calculate fragmentation spectra, which improved the recognition of lipids. Furthermore, the scoring function was improved by adding metadata, such as citation counts, to score those compounds on a higher rank, which were mentioned more often in articles [87]. CFM-ID also can be used to predict the

chemical class of compounds. This was proved in the article by applying this tool to a dataset of 208 MS/MS spectra. Out of this dataset, the chemical classes of 168 compounds were correctly assigned [87].

Ruttkies et al. developed a new version of MetFrag and compared this tool with CFM-ID [86]. In MetFrag, the first step of fragmentation is to generate, according to its accurate mass, a suspect list by querying public databases, such as PubChem [97]. Several filters are then applied to narrow down the list of suspects by applying, for example, InChIKey filtering, which only uses the skeleton of the compound being coded in the first block of InChIKey, to avoid calculating the spectra of isomers, which cannot be separated by MS/MS spectra [86]. Other filtering steps are substructure restrictions and element restrictions [86]. Furthermore, the compounds are scored, according to the number of patents and the number of references in which the compound was mentioned [86]. This approach was tested by recording 473 spectra of 359 reference standards, measured with different collision energies, which resulted in 105 top-1-ranked spectra [86]. In addition, the authors compared MetFrag with CFM-ID. The result for this comparison showed that CFM-ID ranked more compounds as top one than MetFrag and also had a better overall performance. Hence, these improved results had a big disadvantage, because CFM-ID needed more computation time, with 150 min per query in comparison to 54 s with MetFrag [86]. The addition of retention time information to the scoring process even led to the improvement to 87% of top-1-ranked compounds for the earlier mentioned datasets [86]. This demonstrates the possible improvements which could be achieved by in silico fragmentation, especially if no reference data is available, just by taking different sources into account.

For the improvement of the MetFrag workflow, Schymanski et al. recently published an article in which it was described that the initial query of possible compounds in PubChem [97] could be improved by condensing down more than 100 M compounds to just 360 k of the most important compounds [98]. This approach increased the top-1-ranked compounds of a dataset containing 1336 chemicals from 58% to 70% [98]. Furthermore, da Silva et al. developed a machine-learning approach called "Network Annotation Propagation" (NAP), which uses machine learning to detect similarities in MS/MS spectra [99]. This approach aims to improve the annotation of in silico fragmentation by assigning possible structure features found in other spectra, which could fill the gaps of this process [99]. The authors benchmarked this approach by comparing it with MetFrag, which resulted in an overall increase in rank of up to 18% [99].

Blaženović et al. conducted a comparative study of four different in silico fragmentation tools, MetFrag [86], CFM-ID [87], MAGMa+ [100] and MS-Finder [101], using 312 spectra as the training set and 208 spectra as the test set of the Casmi 2016 contest [102,103]. With a 5-ppm search window, a suspect list for both sets was acquired from ChemSpider [104] to have a common base for all four tools. The training set was used to optimize the parameters of these tools [102]. In the first study, MetFrag performed best with 17% of top-1-ranked compounds of the training set [102]. A combination of MetFrag and CFM-ID improved this result to 22%, and by boosting the query with databases and metadata, a rate of 93% of top-1-ranked compounds was achieved. By using just one in silico fragmentation tool with database boosting, a rate of 86% of correct top-1-ranked compounds was achieved with CFM-ID for the test set, whereas MetFrag performed best without boosting, with a rate of 25% [102]. These results show that a combination of in silico fragmentation, additional metadata and database search represents a promising workflow for non-target screening.

Such an approach was published by Gerlich et al. [105]. The tool in this article is called "MetFusion", which combines in silico fragmentation with MetFrag and database search with MassBank [45]. In this workflow, the acquired spectra are passed to MetFrag and MassBank and the hereby received compound candidates are compared with each other by their Tanimoto similarity [82,105]. By doing this, a combined score is calculated for the suspects and the results of MetFrag are reranked. This workflow was evaluated by applying it on a dataset of 1099 spectra, using a similarity filter of 0.9. These spectra were,

in a first step, queried only with MetFrag, which resulted in a median rank of 28 of the correct compounds [105]. Using MetFusion, the median rank improved by far, being seven.

This proves that a combination of spectral databases and in silico fragmentation could be a promising approach, especially if the tools become more precise with further research.

5.5. Annotation with Mass Spectral Databases

As described in the previous chapters, mass spectral databases are crucial in non-target screening to obtain high true positive rates. For those databases, tandem in time instruments, such as Orbitrap and QTOF, are most suitable for the creation of databases, due to their higher resolution and sensitivity [9,34]. Because of the missing standardization of MS instruments, a common framework for the creation of mass spectral databases has to be found to achieve reliable quality of inter-laboratory data for mass spectral repositories [9]. Recently, Oberacher et al. proposed such a framework [106]. For use in non-target screening, according to the authors, mass spectral databases should fulfil the following criteria. First, data should be acquired with high-resolution instrumentation, a minimum MS/MS resolution of 10,000 and a mass error below 10 ppm. Then, the ionization methods should be ESI, APCI or atmospheric pressure photoionization (APPI), and the isolation of the precursor ion should be chosen to be as narrow as possible [106]. This implies that the sample has to be ionizable by these methods in either positive or negative mode. The fragmentation mode method should also be either HCD or CID. Due to the small masses of metabolites, the mass range should start at least at an m/z of 50, and it is preferred to be noted in the metadata of the library [106]. To cover the breakdown curve of the fragmentation process, spectra with multiple collision energies (at least three) should be included in the databases. In addition, a proper curation and expert review is advised for databases [106]. A description of these proposals and of the proposal for the database acquisition can be found in the article of Oberacher et al. [106].

In the previous chapters, several mass spectral databases have already been mentioned, and a deeper description of those repositories can be found in the following references [85,107,108].

An extensive description of the creation of a database was published by Shahaf et al. [109,110]. The herewith created database "WEIZMASS" includes 3039 unique compounds acquired with a QTOF instrument in positive and negative mode, but in contrast to the aforementioned proposal, the authors used a ramp of collision energies from 10 to 30 eV in positive mode, and 15 to 35 eV in negative mode, respectively [110]. The standards were pooled in batches of 20 compounds based on their accurate masses and expected RT by comparing their log D [109,110]. For the extraction and curation process of the spectra, the authors used XCMS and Camera. This process was automated to achieve a high throughput [110]. Furthermore, a matching algorithm was developed called "MatchWeiz" and the authors evaluated its capacities with a validation set of 100 standard compounds [109].

The creation of another database, the EMBL-MCF database, was recently published by Phapale et al. [111]. This database includes 1611 spectra from 435 standard compounds acquired on an Orbitrap instrument with three different collision energies [111]. In addition to MS/MS spectra, the retention times and settings of six different chromatographic systems used were included in the library. For metabolites, the authors used HILIC-LC-MS and for lipids, RP-LC-MS was used [111].

A general description of practices and pitfalls in searching a mass spectral library can be found in the article by Stein [5].

The creation of search algorithms is a crucial step because the manual annotation of compounds with mass spectral data is laborious. Oberacher et al. published a well-established search algorithm MSforID [112] which has been used in several publications [24,48,106,112–115] and can be used for the annotation of all compounds, which are included in the database. The inputs needed by the algorithm are the precursor ion, the accurate masses of the fragments and their relative intensities [112]. Other inputs for this algorithm, which can be chosen by the user, are the mass difference and an intensity threshold. The precise formula of this algorithm can be found in refer-

ence [112]. This algorithm was evaluated against the search algorithm of NIST MS Search by Oberacher et al. [115]. The main difference between both algorithms is that MSforID uses the average similarity of all reference spectra for a certain compound with the acquired spectra, whereas NIST MS Search uses the similarity of a specific reference spectrum with the acquired spectrum [115]. In this study, the authors concluded that both approaches resulted in a similar sensitivity [115].

In chapter 4, the data processing software XCMS was mentioned. In addition to data processing, XCMS Online [62] also implemented statistical tools for finding features with the highest significance and sample clustering [116]. Furthermore, the METLIN database was integrated in the XCMS Online software for MS/MS matching using a cosine similarity score [116]. Benton et al. described a workflow for autonomous metabolite annotation [116]. To validate this workflow, the authors compared the MS only mode with *Desulfovibrio vulgaris Hildenborough* biofilm samples acquired in DIA mode [116]. The results show that 29% of all features were detected by both methods; however, 49% were detected by MS only due to the higher signal intensity and peak definition [116]. This difference occurs because of the longer acquisition time in MS only mode. In contrast to MS only, the MS/MS mode acquires time MS only data in 25% of the cycle and fragmentation spectra in the rest of the cycle time span. In addition to the feature detection validation, the authors conducted a search in the METLIN database for the biofilm sample using their search algorithm, which resulted in 67 matches, representing 20% of the features [116]. For 36 of these metabolites, the KEGG ID was found, which was used for biochemical pathway mapping. This shows that non-target screening of metabolites can result in a more extensive gain of knowledge of the sample, especially if more spectra are included in publicly available databases.

As described in the previous chapters, mass spectral databases of reference standards are a key tool to enable researchers to conduct non-target screening and to promote gaining more insight into the metabolic composition of samples if the amount of spectral data in databases grows due to collaborative efforts.

6. Applications

Non-target screening can be applied in various fields of phyto-analysis. Several reviews were published describing workflows for non-target screening, and the most important steps are also mentioned in this review. These workflows can be found in references [34,108,117–121]. Kalogiouri et al. gave an overview of different efforts for the analysis of olive oils [34], in particular, distinguishing differences between extra virgin and virgin olive oil, which corresponds to its quality, by non-target metabolite profiling and the additional use of chemometrical tools [34]. In addition, some articles were mentioned in which the origins of olive oils were discriminated with non-target screening [34]. Another application of quality control with non-target screening in phyto samples is the determination of toxic compounds, such as natural toxins or pesticides, which could be facilitated by applying non-target screening routinely because of its high throughput capacity and relatively simple sample preparation, especially if more appropriate high-quality spectra are available in public repositories. Righetti et al. described in their review different methods for evaluating contaminations of mycotoxins originating from fungi infestation in crops [117]. Carlier et al. identified and quantified toxic compounds of sea mango [122]. An additional approach was published by Pérez-Ortega et al. [120] by using an in-house database for the annotation of contaminants, such as mycotoxins and pesticides in food samples.

The identification of pesticide residues and other contaminants in phyto samples is an important task due to their possible toxic effects, which is why these and other compounds are regulated worldwide [119,120,123]. Interesting studies were published which investigated the pesticide contamination of alcoholic beverages, such as beer and wine [2,124–126]. Bolaños et al. studied the presence of pesticides in 5 beer samples and 15 wine samples by acquiring data with a UHPLC–MS/MS system [2]. Due to matrix effects in wine, the wine samples were diluted. The results showed that no pesticides were present

in beer samples, by several were found in several wine samples [2]. This result corresponds to the study of Inoue et al. [124]. This study determined the number of pesticides present throughout the different steps of beer brewing by using LC-MS/MS. For this purpose, the authors spiked ground malt with more than 300 pesticides and brewed beer with this contaminated malt. The results showed that most pesticides were reduced in the wort and adsorbed onto the grain after meshing. At the end of the brewing process, only pesticides having a log P below 2 were found in the finished beer. Apart from beer, several articles on wine analysis have also been published. A review which describes different studies of untargeted wine analysis was published by Pinu et al. [127]. Ruocco et al. were able to determine the vintage and color of German and Italian wines by conducting metabolite profiling [128]. In addition, Arbulu et al. identified 411 metabolites of Graciano red wine and were able to differentiate Tempranillo and Graciano wines from each other using 15 metabolites as biomarkers [129]. Further, the authors created a database containing 2080 oenological compounds [129]. Arapitsas et al. could differentiate six different wine cultivars by conducting chemometrical marker detection and a principal component analysis [130]. Diaz et al. could differentiate three wines of protected denominations of Spanish origin using their metabolic profile [131]. A similar approach was described by Li et al. for the differentiation of five truffle species by their metabolic profiles [132].

Most of the non-target approaches concerning phyto-analysis had the goal of identify phenolic compounds [28,133–138]. The majority of authors of these articles performed a manual annotation of the compounds with data in the literature or in spectral databases [133–137]. Lin et al. conducted a profiling of oligomeric proanthocyanidins by comparing the acquired MS/MS spectra with computed fragment spectra [139]. A metabolite profiling of *Rhus coriaria* (Sumac) annotating metabolites with the literature was conducted by Abu-Reidah et al. [140]. Furthermore, Regazzoni et al. profiled gallotannins and flavonoids with mass spectral databases [141].

El Sayed et al. conducted a characterization of aloe vera species by comparing acquired MS/MS spectra with the literature [138]. This plant is also widely used in cosmetics, and was screened for illicit contents by Meng et al. [142]. In this study, 123 cosmetic samples were screened with Orbitrap using an in-house mass spectral database for the annotation of compounds in these samples [142].

7. Shortcomings

The above-mentioned studies show that non-target screening can be used for various applications. It is particularly very promising for quality control of food and cosmetics or screening for prohibited contents, due to its high throughput and sensitivity. Another application is the metabolite profiling of phyto samples, which could lead to the identification of new unknown metabolites in plants. However, the bottleneck of this approach is the low coverage of metabolites in the chemical space of public spectral databases, which is why joint efforts are important to increase the number of high-quality spectra of reference standards to increase the coverage of the chemical space by those databases. In addition, these techniques are used to determine the concentration of the annotated metabolites.

8. Future Perspectives

In the previous chapters, different tools for application in non-target screening in phyto-analysis were described. To summarize future perspective approaches of these tools, it can be said that machine learning has the power to facilitate the annotation process in non-target screening. In particular, in the fields of retention time prediction, in silico fragmentation and search algorithms for mass spectral databases, more research and the use of computational developments would make these tools more reliable and would lead to less false positive identifications, which would increase the sensitivity of non-target screening. Despite qualitative identification, another possible use of machine learning in non-target screening might be the quantification of compounds. This could especially be useful because the

concentration of metabolites is important in phyto-analysis. Recently, Kruve published a feature on quantification techniques in non-target screening (Figure 8) [143].
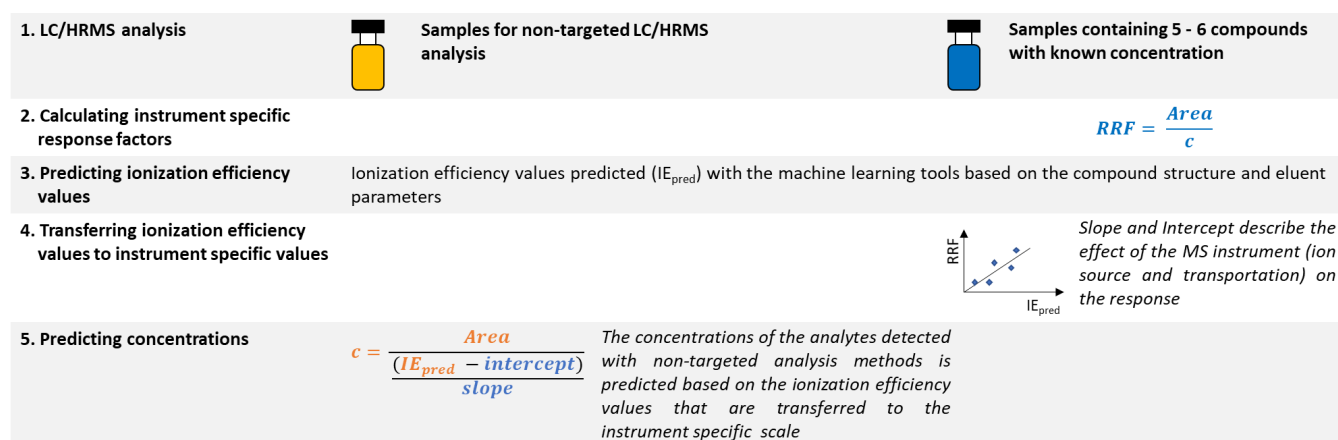

Figure 8. Workflow for using predicted ionization efficiencies for the quantification of compounds discovered in non-targeted analysis, accounting for the structure of the compound, the eluent composition, as well as instrument-specific aspects. [143] Adapted with permission [143] Copyright 2020 American Chemical Society.

Leito et al. proposed the most commonly used parameter, relative ionization efficacy (RIE), for this task [144], which is often used as its logarithmic factor (logIE). Ionization efficiencies are dependent on the amount of compound ionized by ESI and finally detected by the mass spectrometer. The concentration of the compounds can be calculated, using their peak area, by measuring the ionization efficiency. Oss et al. published a scale containing logIEs of 62 compounds and developed a prediction formula based on the compound's pKa and molecular volume [145]. In addition, Abrahamsson et al. [146], Mayhew et al. [147] and Aalizadeh et al. [148] published different machine learning approaches for the quantification of compounds based on ionization efficiencies. Aalizadeh et al. used quantitative structure–property relationship (QSPR) to predict concentrations of phyto metabolites based on logIE [148]. For the prediction of logIE in negative mode, six descriptors were used to build the model. The authors stated that this model showed a fivefold difference compared to the determination of concentrations with reference standards [148]. In addition, Liigand et al. showed that ionization efficiencies are transferable between different instruments [149]. This demonstrates that, in the future, concentrations could be calculated using ionization efficiencies if additional data is acquired to build more precise models. The same can be said for mass spectral databases, which require more data acquired from reference standards to be able to cover the chemical space and annotate more compounds with higher accuracy, which will facilitate and promote non-target screening.

Abbreviations

ANN	Artificial neural networks
APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photoionization
CID	Collision-induced dissociation
CWT	Continuous wavelet transform
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DLM	Deep-learning regression model
DNN	Deep neural networks
DsDA	Data-set-dependent MS/MS
ESI	Electrospray ionization
GC	Gas chromatography
HCD	Higher energy collisional dissociation
HILIC	Hydrophilic interaction chromatography
HPLC	High-performance liquid chromatography
HRMS	High-resolution mass spectrometry
idMS/MS	Indiscriminate MS/MS
IPO	Isotopologue parameter optimization
LC	Liquid chromatography
LLE	Liquid–liquid extraction
logIE	Logarithmic factor of the relative ionization efficacy
MAE	Microwave-assisted extraction
MS	Mass spectrometry
NAP	Network Annotation Propagation
NMR	Nuclear magnetic resonance
NPLC	Normal phase chromatography
PLE	Pressurized liquid extraction
ppm	Parts per million
QC	Quality control
QSPR	Quantitative structure–property relationship
QSRR	Quantitative structure−retention relationships
QuEChERS	Quick, easy, cheap, effective, rugged and safe
QTOF	Quadrupole time-of-flight
RI	Retention time index
RIE	Relative ionization efficacy
ROI	Regions of interest
RP	Reversed phase
SALLE	Salting-out assisted liquid–liquid extraction
SFC	Supercritical fluid chromatography
SFE	Supercritical fluid extraction
SPE	Solid phase extraction
SRM	Single reaction monitoring
SWATH-MS	Sequential windowed acquisition of all theoretical fragment ion mass spectra
TMS	Trimethylsilyl
UAE	Ultrasound-assisted extraction
UHPLC	Ultra-high-performance liquid chromatography
WRTMD	Wiley registry of tandem mass detection
XIC	Extracted ion chromatogram

References

1. New Hope Network. Natural Retail Market Size and Stats | Market Overview. 2020. Available online: https://www.newhope.com/market-data-and-analysis/market-overview-2020-natural-retail-market-size-and-stats (accessed on 13 July 2021).
2. Bolaños, P.P.; Romero-González, R.; Frenich, A.G.; Vidal, J.L.M. Application of hollow fibre liquid phase microextraction for the multiresidue determination of pesticides in alcoholic beverages by ultra-high pressure liquid chromatography coupled to tandem mass spectrometry. *J. Chromatogr. A* **2008**, *1208*, 16–24. [CrossRef] [PubMed]
3. Paar, M. On the History of Austrian Wine Law from 1907 to 1985. *JEHL* **2019**, *10*, 15–25.

4. Daughton, C.G. The Matthew Effect and widely prescribed pharmaceuticals lacking environmental monitoring: Case study of an exposure-assessment vulnerability. *Sci. Total Environ.* **2014**, *466*, 315–325. [CrossRef]

5. Stein, S. Mass spectral reference libraries: An ever-expanding resource for chemical identification. *Anal. Chem.* **2012**, *84*, 7274–7282. [CrossRef] [PubMed]

6. Lachenmeier, D.W.; Humpfer, E.; Fang, F.; Schütz, B.; Dvortsak, P.; Sproll, C.; Spraul, M. NMR-spectroscopy for nontargeted screening and simultaneous quantification of health-relevant compounds in foods: The example of melamine. *J. Agric. Food Chem.* **2009**, *57*, 7194–7199. [CrossRef]

7. Musio, B.; Todisco, S.; Antonicelli, M.; Garino, C.; Arlorio, M.; Mastrorilli, P.; Latronico, M.; Gallo, V. Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production. *Appl. Sci.* **2022**, *12*, 2583. [CrossRef]

8. Remane, D.; Wissenbach, D.K.; Peters, F.T. Recent advances of liquid chromatography-(tandem) mass spectrometry in clinical and forensic toxicology—An update. *Clin. Biochem.* **2016**, *49*, 1051–1071. [CrossRef]

9. Oberacher, H.; Arnhard, K. Compound identification in forensic toxicological analysis with untargeted LC-MS-based techniques. *Bioanalysis* **2015**, *7*, 2825–2840. [CrossRef]

10. de Vijlder, T.; Valkenborg, D.; Lemière, F.; Romijn, E.P.; Laukens, K.; Cuyckens, F. A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation. *Mass Spectrom. Rev.* **2018**, *37*, 607–629. [CrossRef]

11. Milman, B.L.; Zhurkovich, I.K. The chemical space for non-target analysis. *TrAC Trends Anal. Chem.* **2017**, *97*, 179–187. [CrossRef]

12. Milman, B.L.; Zhurkovich, I.K. Mass spectral libraries: A statistical review of the visible use. *TrAC Trends Anal. Chem.* **2016**, *80*, 636–640. [CrossRef]

13. Bade, R.; Causanilles, A.; Emke, E.; Bijlsma, L.; Sancho, J.V.; Hernandez, F.; de Voogt, P. Facilitating high resolution mass spectrometry data processing for screening of environmental water samples: An evaluation of two deconvolution tools. *Sci. Total Environ.* **2016**, *569*, 434–441. [CrossRef] [PubMed]

14. Bader, T.; Schulz, W.; Kümmerer, K.; Winzenbacher, R. LC-HRMS Data Processing Strategy for Reliable Sample Comparison Exemplified by the Assessment of Water Treatment Processes. *Anal. Chem.* **2017**, *89*, 13219–13226. [CrossRef] [PubMed]

15. Hollender, J.; Schymanski, E.L.; Singer, H.P.; Ferguson, P.L. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* **2017**, *51*, 11505–11512. [CrossRef]

16. Schymanski, E.L.; Singer, H.P.; Slobodnik, J.; Ipolyi, I.M.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; et al. Non-target screening with high-resolution mass spectrometry: Critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem.* **2015**, *407*, 6237–6255. [CrossRef]

17. Fischer, K.; Fries, E.; Körner, W.; Schmalz, C.; Zwiener, C. New developments in the trace analysis of organic water pollutants. *Appl. Microbiol. Biotechnol.* **2012**, *94*, 11–28. [CrossRef]

18. Burgard, D.A.; Banta-Green, C.; Field, J.A. Working upstream: How far can you go with sewage-based drug epidemiology? *Environ. Sci. Technol.* **2014**, *48*, 1362–1368. [CrossRef]

19. Urbas, A.; Schoenberger, T.; Corbett, C.; Lippa, K.; Rudolphi, F.; Robien, W. NPS Data Hub: A web-based community driven analytical data repository for new psychoactive substances. *Forensic Chem.* **2018**, *9*, 76–81. [CrossRef]

20. Causanilles, A.; Kinyua, J.; Ruttkies, C.; van Nuijs, A.L.N.; Emke, E.; Covaci, A.; de Voogt, P. Qualitative screening for new psychoactive substances in wastewater collected during a city festival using liquid chromatography coupled to high-resolution mass spectrometry. *Chemosphere* **2017**, *184*, 1186–1193. [CrossRef]

21. Bade, R.; Tscharke, B.J.; White, J.M.; Grant, S.; Mueller, J.F.; O'Brien, J.; Thomas, K.V.; Gerber, C. LC-HRMS suspect screening to show spatial patterns of New Psychoactive Substances use in Australia. *Sci. Total Environ.* **2019**, *650*, 2181–2187. [CrossRef]

22. Bade, R.; White, J.M.; Gerber, C. Qualitative and quantitative temporal analysis of licit and illicit drugs in wastewater in Australia using liquid chromatography coupled to mass spectrometry. *Anal. Bioanal. Chem.* **2018**, *410*, 529–542. [CrossRef] [PubMed]

23. Pasin, D.; Cawley, A.; Bidny, S.; Fu, S. Current applications of high-resolution mass spectrometry for the analysis of new psychoactive substances: A critical review. *Anal. Bioanal. Chem.* **2017**, *409*, 5821–5836. [CrossRef] [PubMed]

24. Reinstadler, V.; Lierheimer, S.; Boettcher, M.; Oberacher, H. A validated workflow for drug detection in oral fluid by non-targeted liquid chromatography-tandem mass spectrometry. *Anal. Bioanal. Chem.* **2019**, *411*, 867–876. [CrossRef] [PubMed]

25. Jorge, T.F.; Rodrigues, J.A.; Caldana, C.; Schmidt, R.; van Dongen, J.T.; Thomas-Oates, J.; António, C. Mass spectrometry-based plant metabolomics: Metabolite responses to abiotic stress. *Mass Spectrom. Rev.* **2016**, *35*, 620–649. [CrossRef]

26. Azwanida, N.N. A Review on the Extraction Methods Use in Medicinal Plants, Principle, Strength and Limitation. *Med Aromat Plants* **2015**, *4*, 1–6. [CrossRef]

27. Martínez Vidal, J.L.; Plaza-Bolaños, P.; Romero-González, R.; Garrido Frenich, A. Determination of pesticide transformation products: A review of extraction and detection methods. *J. Chromatogr. A* **2009**, *1216*, 6767–6788. [CrossRef]

28. Oniszczuk, A.; Olech, M. Optimization of ultrasound-assisted extraction and LC-ESI–MS/MS analysis of phenolic acids from *Brassica oleracea* L. var. sabellica. *Ind. Crops Prod.* **2016**, *83*, 359–363. [CrossRef]

29. Farré, M.; Kantiani, L.; Petrovic, M.; Pérez, S.; Barceló, D. Achievements and future trends in the analysis of emerging organic contaminants in environmental samples by mass spectrometry and bioanalytical techniques. *J. Chromatogr. A* **2012**, *1259*, 86–99. [CrossRef]

30. Mustafa, A.; Turner, C. Pressurized liquid extraction as a green approach in food and herbal plants extraction: A review. *Anal. Chim. Acta* **2011**, *703*, 8–18. [CrossRef]

31. Arumugham, T.; Rambabu, K.; Hasan, S.W.; Show, P.L.; Rinklebe, J.; Banat, F. Supercritical carbon dioxide extraction of plant phytochemicals for biological and environmental applications—A review. *Chemosphere* **2021**, *271*, 129525. [CrossRef]

32. Kole, P.L.; Venkatesh, G.; Kotecha, J.; Sheshala, R. Recent advances in sample preparation techniques for effective bioanalytical methods. *Biomed. Chromatogr.* **2011**, *25*, 199–217. [CrossRef] [PubMed]

33. Bylda, C.; Thiele, R.; Kobold, U.; Volmer, D.A. Recent advances in sample preparation techniques to overcome difficulties encountered during quantitative analysis of small molecules from biofluids using LC-MS/MS. *Analyst* **2014**, *139*, 2265–2276. [CrossRef] [PubMed]

34. Kalogiouri, N.P.; Aalizadeh, R.; Dasenaki, M.E.; Thomaidis, N.S. Application of High Resolution Mass Spectrometric methods coupled with chemometric techniques in olive oil authenticity studies—A review. *Anal. Chim. Acta* **2020**, *1134*, 150–173. [CrossRef] [PubMed]

35. He, S.; Zhang, L.; Bai, S.; Yang, H.; Cui, Z.; Zhang, X.; Li, Y. Advances of molecularly imprinted polymers (MIP) and the application in drug delivery. *Eur. Polym. J.* **2021**, *143*, 110179. [CrossRef]

36. Cajka, T.; Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **2016**, *88*, 524–545. [CrossRef]

37. Dudzik, D.; Barbas-Bernardos, C.; García, A.; Barbas, C. Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *J. Pharm. Biomed. Anal.* **2018**, *147*, 149–173. [CrossRef]

38. Schulze, B.; Jeon, Y.; Kaserzon, S.; Heffernan, A.L.; Dewapriya, P.; O'Brien, J.; Gomez Ramos, M.J.; Ghorbani Gorji, S.; Mueller, J.F.; Thomas, K.V.; et al. An assessment of quality assurance/quality control efforts in high resolution mass spectrometry non-target workflows for analysis of environmental samples. *TrAC Trends Anal. Chem.* **2020**, *133*, 116063. [CrossRef]

39. Niessen, W.M.A.; Manini, P.; Andreoli, R. Matrix effects in quantitative pesticide analysis using liquid chromatography-mass spectrometry. *Mass Spectrom. Rev.* **2006**, *25*, 881–899. [CrossRef]

40. Peters, F.T.; Remane, D. Aspects of matrix effects in applications of liquid chromatography-mass spectrometry to forensic and clinical toxicology—A review. *Anal. Bioanal. Chem.* **2012**, *403*, 2155–2172. [CrossRef]

41. Bahr, U.; Pfenninger, A.; Karas, M.; Stahl, B. High-sensitivity analysis of neutral underivatized oligosaccharides by nanoelectrospray mass spectrometry. *Anal. Chem.* **1997**, *69*, 4530–4535. [CrossRef]

42. Liigand, J.; Laaniste, A.; Kruve, A. pH Effects on Electrospray Ionization Efficiency. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 461–469. [CrossRef] [PubMed]

43. Kruve, A.; Kaupmees, K. Adduct Formation in ESI/MS by Mobile Phase Additives. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 887–894. [CrossRef] [PubMed]

44. Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S.S.; Wohlgemuth, G.; Barupal, D.K.; Showalter, M.R.; Arita, M.; et al. Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **2018**, *37*, 513–532. [CrossRef] [PubMed]

45. MassBank. MassBank|MassBank Europe Mass Spectral DataBase. Available online: https://massbank.eu/MassBank/ (accessed on 30 June 2021).

46. METLIN Database. Available online: https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage (accessed on 30 June 2021).

47. Domingo-Almenara, X.; Guijas, C.; Billings, E.; Montenegro-Burke, J.R.; Uritboonthai, W.; Aisporna, A.E.; Chen, E.; Benton, H.P.; Siuzdak, G. The METLIN small molecule dataset for machine learning-based retention time prediction. *Nat. Commun.* **2019**, *10*, 5811. [CrossRef]

48. Oberacher, H.; Reinstadler, V.; Kreidl, M.; Stravs, M.A.; Hollender, J.; Schymanski, E.L. Annotating Nontargeted LC-HRMS/MS Data with Two Complementary Tandem Mass Spectral Libraries. *Metabolites* **2018**, *9*, 3. [CrossRef]

49. Oberacher, H. *Wiley Registry of Tandem Mass Spectral Data, MS for ID*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2011.

50. Gillet, L.C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteom.* **2012**, *11*, O111.016717. [CrossRef]

51. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **2015**, *12*, 523–526. [CrossRef]

52. Guo, J.; Huan, T. Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography-Mass Spectrometry Based Untargeted Metabolomics. *Anal. Chem.* **2020**, *92*, 8072–8080. [CrossRef]

53. Broeckling, C.D.; Heuberger, A.L.; Prince, J.A.; Ingelsson, E.; Prenni, J.E. Assigning precursor–product ion relationships in indiscriminant MS/MS data from non-targeted metabolite profiling studies. *Metabolomics* **2013**, *9*, 33–43. [CrossRef]

54. Smith, C.A.; Want, E.J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006**, *78*, 779–787. [CrossRef]

55. Kuhl, C.; Tautenhahn, R.; Neumann, S. LC-MS peak annotation and identification with CAMERA. *Anal. Chem.* **2019**, *84*, 1–14.

56. Broeckling, C.D.; Hoyes, E.; Richardson, K.; Brown, J.M.; Prenni, J.E. Comprehensive Tandem-Mass-Spectrometry Coverage of Complex Samples Enabled by Data-Set-Dependent Acquisition. *Anal. Chem.* **2018**, *90*, 8020–8027. [CrossRef] [PubMed]

57. Katajamaa, M.; Oresic, M. Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A* **2007**, *1158*, 318–328. [CrossRef] [PubMed]

58. Domingo-Almenara, X.; Montenegro-Burke, J.R.; Benton, H.P.; Siuzdak, G. Annotation: A Computational Solution for Streamlining Metabolomics Analysis. *Anal. Chem.* **2018**, *90*, 480–489. [CrossRef] [PubMed]

59. Tautenhahn, R.; Böttcher, C.; Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinform.* **2008**, *9*, 504. [CrossRef] [PubMed]

60. Katajamaa, M.; Oresic, M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinform.* **2005**, *6*, 179. [CrossRef]

61. McLean, C.; Kujawinski, E.B. AutoTuner: High Fidelity and Robust Parameter Selection for Metabolomics Data Processing. *Anal. Chem.* **2020**, *92*, 5724–5732. [CrossRef]

62. XCMS Metabolomic and Lipidomic Platform. Available online: https://xcmsonline.scripps.edu/landing_page.php?pgcontent=mainPage (accessed on 30 June 2021).

63. Rafiei, A.; Sleno, L. Comparison of peak-picking workflows for untargeted liquid chromatography/high-resolution mass spectrometry metabolomics data analysis. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 119–127. [CrossRef]

64. MATLAB-MathWorks. Available online: https://www.mathworks.com/products/matlab.html (accessed on 30 June 2021).

65. Conley, C.J.; Smith, R.; Torgrip, R.J.O.; Taylor, R.M.; Tautenhahn, R.; Prince, J.T. Massifquant: Open-source Kalman filter-based XC-MS isotope trace feature detection. *Bioinformatics* **2014**, *30*, 2636–2643. [CrossRef]

66. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [CrossRef]

67. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* **2010**, *11*, 395. [CrossRef] [PubMed]

68. Smith, R.; Tostengard, A.R. Quantitative Evaluation of Ion Chromatogram Extraction Algorithms. *J. Proteome Res.* **2020**, *19*, 1953–1964. [CrossRef] [PubMed]

69. Loos, M.J. Mining of High-Resolution Mass Spectrometry Data to Monitor Organic Pollutant Dynamics in Aquatic Systems. Ph.D. Thesis, ETH Zurich, Zurich, Switzerland, 2015. [CrossRef]

70. Hohrenk, L.L.; Itzel, F.; Baetz, N.; Tuerk, J.; Vosough, M.; Schmidt, T.C. Comparison of Software Tools for Liquid Chromatography-High-Resolution Mass Spectrometry Data Processing in Nontarget Screening of Environmental Samples. *Anal. Chem.* **2020**, *92*, 1898–1907. [CrossRef] [PubMed]

71. Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinform.* **2007**, *8*, 105. [CrossRef]

72. Ljoncheva, M.; Stepišnik, T.; Džeroski, S.; Kosjek, T. Cheminformatics in MS-based environmental exposomics: Current achievements and future directions. *Trends Environ. Anal. Chem.* **2020**, *28*, e00099. [CrossRef]

73. Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I. InChI-the worldwide chemical structure identifier standard. *J. Cheminform.* **2013**, *5*, 7. [CrossRef]

74. Heller, S.R.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. *J. Cheminform.* **2015**, *7*, 23. [CrossRef]

75. InChI Web Service. Available online: https://www.chemspider.com/InChI.asmx (accessed on 1 July 2021).

76. CompTox Chemicals Dashboard. Available online: https://comptox.epa.gov/dashboard/dsstoxdb/batch_search (accessed on 1 July 2021).

77. Schymanski, E.L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H.P.; Hollender, J. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* **2014**, *48*, 2097–2098. [CrossRef]

78. Creek, D.J.; Jankevics, A.; Breitling, R.; Watson, D.G.; Barrett, M.P.; Burgess, K.E.V. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: Improved metabolite identification by retention time prediction. *Anal. Chem.* **2011**, *83*, 8703–8710. [CrossRef]

79. Aalizadeh, R.; Nika, M.-C.; Thomaidis, N.S. Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants. *J. Hazard. Mater.* **2019**, *363*, 277–285. [CrossRef]

80. Stanstrup, J.; Neumann, S.; Vrhovšek, U. PredRet: Prediction of retention time by direct mapping between multiple chromatographic systems. *Anal. Chem.* **2015**, *87*, 9421–9428. [CrossRef] [PubMed]

81. Hall, L.M.; Hill, D.W.; Bugden, K.; Cawley, S.; Hall, L.H.; Chen, M.-H.; Grant, D.F. Development of a Reverse Phase HPLC Retention Index Model for Nontargeted Metabolomics Using Synthetic Compounds. *J. Chem. Inf. Model.* **2018**, *58*, 591–604. [CrossRef] [PubMed]

82. Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **2015**, *7*, 20. [CrossRef] [PubMed]

83. Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8*, 31. [CrossRef]

84. Nguyen, D.H.; Nguyen, C.H.; Mamitsuka, H. Recent advances and prospects of computational methods for metabolite identification: A review with emphasis on machine learning approaches. *Brief. Bioinform.* **2019**, *20*, 2028–2043. [CrossRef] [PubMed]

85. Moumbock, A.F.A.; Ntie-Kang, F.; Akone, S.H.; Li, J.; Gao, M.; Telukunta, K.K.; Günther, S. An overview of tools, software, and methods for natural product fragment and mass spectral analysis. *Phys. Sci. Rev.* **2019**, *4*, 20180126. [CrossRef]

86. Ruttkies, C.; Schymanski, E.L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. *J. Cheminform.* **2016**, *8*, 3. [CrossRef]

87. Djoumbou-Feunang, Y.; Pon, A.; Karu, N.; Zheng, J.; Li, C.; Arndt, D.; Gautam, M.; Allen, F.; Wishart, D.S. CFM-ID 3.0: Significantly Improved ESI-MS/MS Prediction and Compound Identification. *Metabolites* **2019**, *9*, 72. [CrossRef]

88. O'Shea, K.; Misra, B.B. Software tools, databases and resources in metabolomics: Updates from 2018 to 2019. *Metabolomics* **2020**, *16*, 36. [CrossRef]

89. Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A.A.; Melnik, A.V.; Meusel, M.; Dorrestein, P.C.; Rousu, J.; Böcker, S. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **2019**, *16*, 299–302. [CrossRef]

90. Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI: FingerID. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12580–12585. [CrossRef] [PubMed]

91. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [CrossRef] [PubMed]

92. MassBank of North America (MoNA). Available online: https://mona.fiehnlab.ucdavis.edu/ (accessed on 6 July 2021).

93. PhytoHub. Available online: https://phytohub.eu/ (accessed on 6 July 2021).

94. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res* **2018**, *46*, D1074–D1082. [CrossRef]

95. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res* **2018**, *46*, D608–D617. [CrossRef] [PubMed]

96. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [CrossRef]

97. PubChem. Available online: https://pubchem.ncbi.nlm.nih.gov/ (accessed on 6 July 2021).

98. Schymanski, E.L.; Kondić, T.; Neumann, S.; Thiessen, P.A.; Zhang, J.; Bolton, E.E. Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. *J. Cheminform.* **2021**, *13*, 19. [CrossRef] [PubMed]

99. Da Silva, R.R.; Wang, M.; Nothias, L.-F.; van der Hooft, J.J.J.; Caraballo-Rodríguez, A.M.; Fox, E.; Balunas, M.J.; Klassen, J.L.; Lopes, N.P.; Dorrestein, P.C. Propagating annotations of molecular networks using in silico fragmentation. *PLoS Comput. Biol.* **2018**, *14*, e1006089. [CrossRef]

100. Verdegem, D.; Lambrechts, D.; Carmeliet, P.; Ghesquière, B. Improved metabolite identification with MIDAS and MAGMa through MS/MS spectral dataset-driven parameter optimization. *Metabolomics* **2016**, *12*, 98. [CrossRef]

101. Tsugawa, H.; Kind, T.; Nakabayashi, R.; Yukihira, D.; Tanaka, W.; Cajka, T.; Saito, K.; Fiehn, O.; Arita, M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* **2016**, *88*, 7946–7958. [CrossRef]

102. Blaženović, I.; Kind, T.; Torbašinović, H.; Obrenović, S.; Mehta, S.S.; Tsugawa, H.; Wermuth, T.; Schauer, N.; Jahn, M.; Biedendieck, R.; et al. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: Database boosting is needed to achieve 93% accuracy. *J. Cheminform.* **2017**, *9*, 32. [CrossRef]

103. Neumann, Emma Schymanski and Steffen. Critical Assessment of Small Molecule Identification. Available online: http://www.casmi-contest.org/2016/ (accessed on 6 July 2021).

104. ChemSpider. Search and Share Chemistry. Available online: http://www.chemspider.com/Default.aspx (accessed on 6 July 2021).

105. Gerlich, M.; Neumann, S. MetFusion: Integration of compound identification strategies. *J. Mass Spectrom.* **2013**, *48*, 291–298. [CrossRef] [PubMed]

106. Oberacher, H.; Sasse, M.; Antignac, J.-P.; Guitton, Y.; Debrauwer, L.; Jamin, E.L.; Schulze, T.; Krauss, M.; Covaci, A.; Caballero-Casero, N.; et al. A European proposal for quality control and quality assurance of tandem mass spectral libraries. *Env. Sci. Eur.* **2020**, *32*, 43. [CrossRef]

107. Vinaixa, M.; Schymanski, E.L.; Neumann, S.; Navarro, M.; Salek, R.M.; Yanes, O. Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *TrAC Trends Anal. Chem.* **2016**, *78*, 23–35. [CrossRef]

108. Bletsou, A.A.; Jeon, J.; Hollender, J.; Archontaki, E.; Thomaidis, N.S. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. *TrAC Trends Anal. Chem.* **2015**, *66*, 32–44. [CrossRef]

109. Shahaf, N.; Rogachev, I.; Heinig, U.; Meir, S.; Malitsky, S.; Battat, M.; Wyner, H.; Zheng, S.; Wehrens, R.; Aharoni, A. The WEIZMASS spectral library for high-confidence metabolite identification. *Nat. Commun.* **2016**, *7*, 12423. [CrossRef]

110. Shahaf, N.; Aharoni, A.; Rogachev, I. A complete pipeline for generating a high-resolution LC-MS-Based reference mass spectra library. In *Plant Metabolomics*; Humana Press: New York, NY, USA, 2018; pp. 193–206.

111. Phapale, P.; Palmer, A.; Gathungu, R.M.; Kale, D.; Brügger, B.; Alexandrov, T. Public LC-Orbitrap Tandem Mass Spectral Library for Metabolite Identification. *J. Proteome Res.* **2021**, *20*, 2089–2097. [CrossRef]

112. Oberacher, H.; Pavlic, M.; Libiseller, K.; Schubert, B.; Sulyok, M.; Schuhmacher, R.; Csaszar, E.; Köfeler, H.C. On the inter-instrument and the inter-laboratory transferability of a tandem mass spectral reference library: 2. Optimization and characterization of the search algorithm. *J. Mass Spectrom.* **2009**, *44*, 494–502. [CrossRef]

113. Oberacher, H.; Whitley, G.; Berger, B. Evaluation of the sensitivity of the 'Wiley registry of tandem mass spectral data, MSforID' with MS/MS data of the 'NIST/NIH/EPA mass spectral library'. *J. Mass Spectrom.* **2013**, *48*, 487–496. [CrossRef]

114. Oberacher, H.; Weinmann, W.; Dresen, S. Quality evaluation of tandem mass spectral libraries. *Anal. Bioanal. Chem.* **2011**, *400*, 2641–2648. [CrossRef]

115. Oberacher, H.; Whitley, G.; Berger, B.; Weinmann, W. Testing an alternative search algorithm for compound identification with the 'Wiley Registry of Tandem Mass Spectral Data, MSforID'. *J. Mass Spectrom.* **2013**, *48*, 497–504. [CrossRef]

116. Benton, H.P.; Ivanisevic, J.; Mahieu, N.G.; Kurczy, M.E.; Johnson, C.H.; Franco, L.; Rinehart, D.; Valentine, E.; Gowda, H.; Ubhi, B.K.; et al. Autonomous metabolomics for rapid metabolite identification in global profiling. *Anal. Chem.* **2015**, *87*, 884–891. [CrossRef] [PubMed]

117. Righetti, L.; Paglia, G.; Galaverna, G.; Dall'Asta, C. Recent Advances and Future Challenges in Modified Mycotoxin Analysis: Why HRMS Has Become a Key Instrument in Food Contaminant Research. *Toxins* **2016**, *8*, 361. [CrossRef] [PubMed]

118. Chaleckis, R.; Meister, I.; Zhang, P.; Wheelock, C.E. Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics. *Curr. Opin. Biotechnol.* **2019**, *55*, 44–50. [CrossRef] [PubMed]

119. Hernández, F.; Sancho, J.V.; Ibáñez, M.; Grimalt, S. Investigation of pesticide metabolites in food and water by LC-TOF-MS. *TrAC Trends Anal. Chem.* **2008**, *27*, 862–872. [CrossRef]

120. Pérez-Ortega, P.; Lara-Ortega, F.J.; Gilbert-López, B.; Moreno-González, D.; García-Reyes, J.F.; Molina-Díaz, A. Screening of Over 600 Pesticides, Veterinary Drugs, Food-Packaging Contaminants, Mycotoxins, and Other Chemicals in Food by Ultra-High Performance Liquid Chromatography Quadrupole Time-of-Flight Mass Spectrometry (UHPLC-QTOFMS). *Food Anal. Methods* **2017**, *10*, 1216–1244. [CrossRef]

121. Knolhoff, A.M.; Croley, T.R. Non-targeted screening approaches for contaminants and adulterants in food using liquid chromatography hyphenated to high resolution mass spectrometry. *J. Chromatogr. A* **2016**, *1428*, 86–96. [CrossRef]

122. Carlier, J.; Guitton, J.; Bévalot, F.; Fanton, L.; Gaillard, Y. The principal toxic glycosidic steroids in Cerbera manghas L. seeds: Identification of cerberin, neriifolin, tanghinin and deacetyltanghinin by UHPLC-HRMS/MS, quantification by UHPLC-PDA-MS. *J. Chromatogr. B* **2014**, *962*, 1–8. [CrossRef]

123. Kunzelmann, M.; Winter, M.; Åberg, M.; Hellenäs, K.-E.; Rosén, J. Non-targeted analysis of unexpected food contaminants using LC-HRMS. *Anal. Bioanal. Chem.* **2018**, *410*, 5593–5602. [CrossRef]

124. Inoue, T.; Nagatomi, Y.; Suga, K.; Uyama, A.; Mochizuki, N. Fate of pesticides during beer brewing. *J. Agric. Food Chem.* **2011**, *59*, 3857–3868. [CrossRef]

125. Nagatomi, Y.; Yoshioka, T.; Yanagisawa, M.; Uyama, A.; Mochizuki, N. Simultaneous LC-MS/MS analysis of glyphosate, glufosinate, and their metabolic products in beer, barley tea, and their ingredients. *Biosci. Biotechnol. Biochem.* **2013**, *77*, 2218–2221. [CrossRef]

126. Anderson, H.E.; Santos, I.C.; Hildenbrand, Z.L.; Schug, K.A. A review of the analytical methods used for beer ingredient and finished product analysis and quality control. *Anal. Chim. Acta* **2019**, *1085*, 1–20. [CrossRef] [PubMed]

127. Pinu, F. Grape and Wine Metabolomics to Develop New Insights Using Untargeted and Targeted Approaches. *Fermentation* **2018**, *4*, 92. [CrossRef]

128. Ruocco, S.; Perenzoni, D.; Angeli, A.; Stefanini, M.; Rühl, E.; Patz, C.-D.; Mattivi, F.; Rauhut, D.; Vrhovsek, U. Metabolite profiling of wines made from disease-tolerant varieties. *Eur. Food Res. Technol.* **2019**, *245*, 2039–2052. [CrossRef]

129. Arbulu, M.; Sampedro, M.C.; Gómez-Caballero, A.; Goicolea, M.A.; Barrio, R.J. Untargeted metabolomic analysis using liquid chromatography quadrupole time-of-flight mass spectrometry for non-volatile profiling of wines. *Anal. Chim. Acta* **2015**, *858*, 32–41. [CrossRef] [PubMed]

130. Arapitsas, P.; Ugliano, M.; Perenzoni, D.; Angeli, A.; Pangrazzi, P.; Mattivi, F. Wine metabolomics reveals new sulfonated products in bottled white wines, promoted by small amounts of oxygen. *J. Chromatogr. A* **2016**, *1429*, 155–165. [CrossRef] [PubMed]

131. Díaz, R.; Gallart-Ayala, H.; Sancho, J.V.; Nuñez, O.; Zamora, T.; Martins, C.P.B.; Hernández, F.; Hernández-Cassou, S.; Saurina, J.; Checa, A. Told through the wine: A liquid chromatography-mass spectrometry interplatform comparison reveals the influence of the global approach on the final annotated metabolites in non-targeted metabolomics. *J. Chromatogr. A* **2016**, *1433*, 90–97. [CrossRef]

132. Li, X.; Zhang, X.; Ye, L.; Kang, Z.; Jia, D.; Yang, L.; Zhang, B. LC-MS-Based Metabolomic Approach Revealed the Significantly Different Metabolic Profiles of Five Commercial Truffle Species. *Front. Microbiol.* **2019**, *10*, 2227. [CrossRef]

133. Spínola, V.; Pinto, J.; Castilho, P.C. Identification and quantification of phenolic compounds of selected fruits from Madeira Island by HPLC-DAD-ESI-MS(n) and screening for their antioxidant activity. *Food Chem.* **2015**, *173*, 14–30. [CrossRef]

134. Kolniak-Ostek, J. Identification and quantification of polyphenolic compounds in ten pear cultivars by UPLC-PDA-Q/TOF-MS. *J. Food Compos. Anal.* **2016**, *49*, 65–77. [CrossRef]

135. Simirgiotis, M.J.; Bórquez, J.; Schmeda-Hirschmann, G. Antioxidant capacity, polyphenolic content and tandem HPLC-DAD-ESI/MS profiling of phenolic compounds from the South American berries Luma apiculata and L. chequén. *Food Chem.* **2013**, *139*, 289–299. [CrossRef]

136. Bastos, K.X.; Dias, C.N.; Nascimento, Y.M.; Da Silva, M.S.; Langassner, S.M.Z.; Wessjohann, L.A.; Tavares, J.F. Identification of Phenolic Compounds from Hancornia speciosa (Apocynaceae) Leaves by UHPLC Orbitrap-HRMS. *Molecules* **2017**, *22*, 143. [CrossRef] [PubMed]

137. Bertin, R.L.; Gonzaga, L.V.; Da Borges, G.S.C.; Azevedo, M.S.; Maltez, H.F.; Heller, M.; Micke, G.A.; Tavares, L.B.B.; Fett, R. Nutrient composition and, identification/quantification of major phenolic compounds in Sarcocornia ambigua (Amaranthaceae) using HPLC–ESI-MS/MS. *Food Res. Int.* **2014**, *55*, 404–411. [CrossRef]

138. El Sayed, A.M.; Ezzat, S.M.; El Naggar, M.M.; El Hawary, S.S. In vivo diabetic wound healing effect and HPLC–DAD–ESI–MS/MS profiling of the methanol extracts of eight Aloe species. *Rev. Bras. Farmacogn.* **2016**, *26*, 352–362. [CrossRef]

139. Lin, L.-Z.; Sun, J.; Chen, P.; Monagas, M.J.; Harnly, J.M. UHPLC-PDA-ESI/HRMSn profiling method to identify and quantify oligomeric proanthocyanidins in plant products. *J. Agric. Food Chem.* **2014**, *62*, 9387–9400. [CrossRef]

140. Abu-Reidah, I.M.; Ali-Shtayeh, M.S.; Jamous, R.M.; Arráez-Román, D.; Segura-Carretero, A. HPLC-DAD-ESI-MS/MS screening of bioactive components from Rhus coriaria L. (Sumac) fruits. *Food Chem.* **2015**, *166*, 179–191. [CrossRef] [PubMed]

141. Regazzoni, L.; Arlandini, E.; Garzon, D.; Santagati, N.A.; Beretta, G.; Maffei Facino, R. A rapid profiling of gallotannins and flavonoids of the aqueous extract of Rhus coriaria L. by flow injection analysis with high-resolution mass spectrometry assisted with database searching. *J. Pharm. Biomed. Anal.* **2013**, *72*, 202–207. [CrossRef] [PubMed]

142. Meng, X.; Bai, H.; Guo, T.; Niu, Z.; Ma, Q. Broad screening of illicit ingredients in cosmetics using ultra-high-performance liquid chromatography-hybrid quadrupole-Orbitrap mass spectrometry with customized accurate-mass database and mass spectral library. *J. Chromatogr. A* **2017**, *1528*, 61–74. [CrossRef]

143. Kruve, A. Strategies for Drawing Quantitative Conclusions from Nontargeted Liquid Chromatography-High-Resolution Mass Spectrometry Analysis. *Anal. Chem.* **2020**, *92*, 4691–4699. [CrossRef]

144. Leito, I.; Herodes, K.; Huopolainen, M.; Virro, K.; Künnapas, A.; Kruve, A.; Tanner, R. Towards the electrospray ionization mass spectrometry ionization efficiency scale of organic compounds. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 379–384. [CrossRef]

145. Oss, M.; Kruve, A.; Herodes, K.; Leito, I. Electrospray ionization efficiency scale of organic compounds. *Anal. Chem.* **2010**, *82*, 2865–2872. [CrossRef]

146. Panagopoulos Abrahamsson, D.; Park, J.-S.; Singh, R.R.; Sirota, M.; Woodruff, T.J. Applications of Machine Learning to In Silico Quantification of Chemicals without Analytical Standards. *J. Chem. Inf. Model.* **2020**, *60*, 2718–2727. [CrossRef] [PubMed]

147. Mayhew, A.W.; Topping, D.O.; Hamilton, J.F. New Approach Combining Molecular Fingerprints and Machine Learning to Estimate Relative Ionization Efficiency in Electrospray Ionization. *ACS Omega* **2020**, *5*, 9510–9516. [CrossRef] [PubMed]

148. Aalizadeh, R.; Panara, A.; Thomaidis, N.S. Development and Application of a Novel Semi-quantification Approach in LC-QToF-MS Analysis of Natural Products. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1412–1423. [CrossRef] [PubMed]

149. Liigand, J.; Kruve, A.; Liigand, P.; Laaniste, A.; Girod, M.; Antoine, R.; Leito, I. Transferability of the electrospray ionization efficiency scale between different instruments. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1923–1930. [CrossRef] [PubMed]